

# Questions We Have

**An Introduction to Philosophy**

By Davis A. Smith

# **Questions We Have**

## **An Introduction to Philosophy**

*By* **Davis A. Smith**

Questions We Have by Davis Smith is, for the most part, an original introduction to Philosophy textbook. It consists of eleven (11) modules (sections or parts) which each handle a different philosophical question a person may have wondered about. Aside from the in depth discussions of the various topics which Davis Smith provides, there are 10 primary source works which are reprinted into this text. The questions discussed span from the abstract, such as the purpose of Philosophy, to the concrete, like our moral duties to other people and animals. The primary sources used are, in order:

- 1 Plato's Apology
- 2 Death by Thomas Nagel
- 3 The Mind-Body Problem by Tim Crane
- 4 Minds, Brains and Programs by John Searle
- 5 Are Libertarianism and Physicalism Compatible? by Davis Smith
- 6 Are Dualism and Libertarianism Compatible? by Davis Smith
- 7 Meditations on First Philosophy by René Descartes
- 8 The Challenge of Cultural Relativism by James Rachels
- 9 The Moral Status of Bloodbending by Davis Smith
- 10 The Moral and Legal Status of Abortion by Mary Warren
- 11 I Was Once a Fetus by Alexander Pruss

The two papers by Davis Smith concerning Libertarianism will be merged into one with future editions of this textbook.

This work is licensed under a [Creative Commons Attribution 4.0](#) license. You are free to copy and redistribute the material in any medium or format, and remix, transform, and build

upon the material for any purpose, even commercially, under the following terms:

- ▶ You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- ▶ You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

# *Contents*

|                                                      |            |
|------------------------------------------------------|------------|
| Preface: Why Study Philosophy?                       | vii        |
| <b>I      What is Philosophy?</b>                    | <b>1</b>   |
| <hr/>                                                |            |
| Part 1: What Philosophy Is and Isn't                 | 2          |
| Plato's Apology                                      | 40         |
| Part 2: Life and Times of Socrates                   | 67         |
| <b>II      Is Death Bad For The Person Who Died?</b> | <b>82</b>  |
| <hr/>                                                |            |
| Death By Thomas Nagel                                | 83         |
| Part 3: How Does Nagel Define Death?                 | 94         |
| <b>III      What is The Mind? What is the Body?</b>  | <b>101</b> |
| <hr/>                                                |            |
| The Mind-Body Problem by Tim Crane                   | 102        |
| Part 4: The Mind-Body Problem                        | 108        |
| Minds, Brains and Programs by John Searle            | 121        |
| Part 5: Can Machines Think?                          | 150        |

|             |                                                           |            |
|-------------|-----------------------------------------------------------|------------|
| <b>IV</b>   | <b>Do We Have Free Will? What is it?</b>                  | <b>166</b> |
| <hr/>       |                                                           |            |
| Part 6:     | The Free Will Debate                                      | 167        |
| Part 7:     | Determinism, Incompatibilism, and Hard Determinism        | 172        |
| Part 8:     | What If We Denied Determinism?                            | 179        |
| Part 9:     | Compatibilism and Soft Determinism                        | 182        |
|             | Are Libertarianism and Physicalism Compatible?            | 187        |
|             | Are Dualism and Libertarianism Compatible?                | 200        |
| <br>        |                                                           |            |
| <b>V</b>    | <b>Does God Exist? If so, why is there evil?</b>          | <b>213</b> |
| <hr/>       |                                                           |            |
| Part 10:    | Arguments for the existence of God                        | 214        |
| Part 11:    | An Argument Against the Existence of God                  | 236        |
| <br>        |                                                           |            |
| <b>VI</b>   | <b>What is Knowledge? How Do We Know Things?</b>          | <b>253</b> |
| <hr/>       |                                                           |            |
|             | Meditations on First Philosophy by René Descartes         | 254        |
| Part 12:    | René Descartes, Life, Times, and Meditations              | 287        |
| Part 13:    | An Overview of Descartes' Meditations on First Philosophy | 291        |
| Part 14:    | Knowledge and Justification                               | 304        |
| <br>        |                                                           |            |
| <b>VII</b>  | <b>Is Morality Culturally Relative?</b>                   | <b>313</b> |
| <hr/>       |                                                           |            |
|             | The Challenge of Cultural Relativism by James Rachels     | 314        |
| Part 15:    | What is Moral Relativism?                                 | 332        |
| Part 16:    | So, If Moral Relativism is Wrong, What Next?              | 355        |
| <br>        |                                                           |            |
| <b>VIII</b> | <b>How Should I Act?</b>                                  | <b>367</b> |
| <hr/>       |                                                           |            |

|                                                 |     |
|-------------------------------------------------|-----|
| Part 17: Normative Ethics                       | 368 |
| Part 18: Utilitarianism                         | 372 |
| Part 19: Kantianism                             | 386 |
| The Moral Status of Bloodbending by Davis Smith | 397 |

## IX What About Real World Cases? 413

|                                                       |     |
|-------------------------------------------------------|-----|
| The Moral and Legal Status of Abortion by Mary Warren | 414 |
| Part 20:What is an Abortion?                          | 438 |
| Part 21: The Abortion Debate (Pro-Choice)             | 445 |
| I Was Once a Fetus by Alexander Pruss                 | 468 |
| Part 22: The Abortion Debate (Pro-Life)               | 484 |

**X** What about duties to animals and future generations? 496

|                                                     |     |
|-----------------------------------------------------|-----|
| Part 23: Vegetarianism/Veganism                     | 497 |
| Part 24: What Are Our Duties to Future Generations? | 515 |

## XI What is justice? What is equality? 529

Part 25: Political Philosophy 530

## Appendices 550

|               |     |
|---------------|-----|
| Common Biases | 550 |
| Glossary      | 573 |
| Bibliography  | 586 |

# *Preface: Why Study Philosophy?*

Philosophy, put lightly, is thinking really hard about something. It is not just your world view or that of another, but it is also the mother of all other subjects. At its core, philosophy is critical thinking taken to a ridiculous extreme (you will read this again later on). The most famous for this is Socrates. Although philosophy has taken this system of reasoning to a formal extreme, in our day to day lives, we think and argue with each other using a less formal, but very similar system. Philosophers ask themselves and others perplexing questions which they want the answer to.

There was a time, long ago, when all subjects were a branch of philosophy, stemming from very basic questions and seeking the answers to them. As time went on, philosophy splintered into various fields which we know today, taking on different standards for evidence and reasoning. Philosophy, as it is today, retains the highest standard for reasoning and argumentation for the stances. But, though the alternative standards, like those found in science, are practically valuable, they can't answer certain kinds of questions, which is where philosophy remains alive and well.



## Questions Philosophers are Prone to Ask:

Here are some examples of questions philosophers are prone to think about and ask (yes, headaches are common). We will be covering the introduction to many of these as the course progresses.

What is the meaning of life? Is there one?

What am I? Am I some soul thing or just matter?

What makes actions right or wrong? Is it relative or is there some hard and fast rule?

Is there such a thing as God?

How does time work? Is there such a thing as time?

What is 'free will'? Do we have it? Does free will even make sense?

What is it to know something? How does that differ from opinion?

We will touch lightly on aspects of the meaning of life in Module II. Souls and other aspects are part of the Mind-Body Problem which is Module III. Whether morality is relative is in Module VII. Moral rectitude is in Module VIII. Arguments for and against the existence of God is in Module V. The Free Will Debate is in Module IV. And, finally, theories of knowledge are in Module VI.

There are other questions which we will ponder as the course progresses, such as the moral status of animals and those who haven't been born yet and various contemporary debates and issues (including the nature of justice and equality). I am willing to wager that at least one question or topic in this class will get you really interested and excited about Philosophy and make you want to take more courses on that subject.

## Why Study Philosophy?

Like I just said, philosophy is critical thinking taken to an extreme... But why should the average person be concerned with

it? When a person wants to be able to lift heavier objects, they train by lifting weights. Training with heavier weights (eventually) makes the smaller amount that they need to carry easier. The mind is no different. Working with an extreme, outlandish, and hard question, knowing how to think about them and the rules for good critical thinking will make debate and reasoning in every day contexts easier.

There are other, potentially more practical or financial, reasons to study Philosophy. Many of you might think that the humanities, like English, Art, and Philosophy, aren't the best way to make a living. Your parents might say something like "O, you want to get a major in Philosophy? Great, there's a new Philosophy factory opening up down the street!" in a sarcastic way. This common thinking is because many don't see the value in Philosophy. For example, in 2015, in his Presidential bid, Marco Rubio regularly said things along the lines of "we need more welders, less philosophers!"<sup>1</sup> He attempted to support this claim by implying or outright stating that welders make more money than those who got degrees in Philosophy.

Both of these claims are incorrect. On average, people with philosophy degrees make more money (yearly) than those who entered the welding trade.<sup>2</sup> When we compare the average pay for people with Philosophy degrees to other majors, we also see that those with this major make more money than any other humanities degree.<sup>3</sup> Holders of this degree make more money than even the average business management degree holder and some STEM fields.

---

<sup>1</sup>Richmond, Emily. "The Reality of the Philosophers vs. Welders Debate," 14 Nov. 2015, [www.theatlantic.com/education/archive/2015/11/philosopher-vs-welders/415890/](http://www.theatlantic.com/education/archive/2015/11/philosopher-vs-welders/415890/).

<sup>2</sup>Sola, Katie. "Sorry, Rubio, But Philosophers Make 78% More Than Welders," 11 Nov. 2015, [www.forbes.com/sites/katiesola/2015/11/11/rubio-welders-philosophers/?sh=2770ff0641b8](http://www.forbes.com/sites/katiesola/2015/11/11/rubio-welders-philosophers/?sh=2770ff0641b8).

<sup>3</sup>Weinberg, Justin. "Philosophy Majors Make More Money Than Majors in any other Humanities Field," 3 Jan. 2019, [www.forbes.com/sites/katiesola/2015/11/11/rubio-welders-philosophers/?sh=2770ff0641b8](http://www.forbes.com/sites/katiesola/2015/11/11/rubio-welders-philosophers/?sh=2770ff0641b8).

But why is this? Getting a degree in Philosophy doesn't make you 'pre-packaged' with the skills to enter a particular job (like going to a trade school), rather a degree in Philosophy and the classes which are entailed by it give you a more important, overarching set of skills, namely creative problem solving. Employers, the world over, are seeking those who can be given a problem and come up with creative or innovated ways of solving it. The courses you take in Philosophy give you significantly harder problems than those you would be presented with in those situations and this makes those problems so much easier for you to handle. Many famous CEOs and company founders got degrees in Philosophy.

# MODULE I

## *What is Philosophy?*

## *Part 1: What*

# *Philosophy Is and Isn't*

Philosophy is not just a way of life, it is not just some outlook which a person takes on. It is also not the product of some really deep thinking. Philosophic thought is not philosophy. One does not have a philosophy. That's not how it works. 'My philosophy on that is...' is like saying 'my science on that is...' What they should say is 'my belief is that...' or 'I have thought long and hard about it and I think...' Philosophy is critical thinking taken to the ridiculous extreme. Since it is, at its core, thinking, one does not simply study philosophy. Philosophy is an activity, it's something that you do. Doing philosophy involves finding a stance on some issue and then coming up with reasons for that stance. This could be a stance that you hold or it could be the stance that your opposition holds in some debate.<sup>4</sup> In order to really truly prove your point, you will want to know why another might believe otherwise and have the skills to show the errors in that reasoning. Engaging in philosophy is how you get the skills to find the errors in the reasoning.

Since philosophic activities tend to go really abstract re-

---

<sup>4</sup>This would be putting yourself in 'their shoes', so to speak, so that you can see where they are coming from and thereby exploit the flaws in their stance.

ally quickly and since the questions which philosophers tend to think about are very difficult to answer, some have claimed that Philosophy is a pointless endeavor. This is, however, not accurate. All fields of study are defined, roughly, in terms of the questions which they are trying to answer or have the ability to answer. The questions in Philosophy tend to be those which other fields can't answer. In those cases, most of the time, the standards are very high, higher than other fields can match. Sometimes, those standards are so high that some claim that they are not possible to answer. Again, this is not accurate. There are answers, even if those answers are difficult to get. Doing philosophy tends to involve the following activities:

| Philosophy Involves               | Example                                                                                                                                                                   |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Resolving Confusion               | Noticing that people misunderstand something because of a word-choice.                                                                                                    |
| Unmasking Assumptions             | Seeing that you/another hold a position for a reason which was not stated.                                                                                                |
| Revealing Presuppositions         | Seeing that there are unexamined/unexplained jumps in reasoning or an aspect of the reasoning which is hidden.                                                            |
| Distinguishing Importance         | Figuring out which things you need to do vs which things you want to do.                                                                                                  |
| Testing Positions                 | Putting yourself in another's shoes to see whether their stance makes sense or to see whether your own stance holds water.                                                |
| Correcting Distortions            | Fixing/identifying misrepresented stances or facts, seeing that you are being misinformed/lied to.                                                                        |
| Looking for Reasons               | Asking yourself why a person would think/do what they are thinking/doing                                                                                                  |
| Examining World-Views             | Seeing whether another's explanation for something or some course of behavior is correct/accurate to the world itself. People can have wrong opinions.                    |
| Questioning Conceptual Frameworks | Seeing whether a particular way of carving the world, a way of thinking about it, is better or worse than another way. We may think that the world is one way, but is it? |

## Part 1.1: How Does One Do Philosophy?

As a kind of inquiry, philosophy is aimed at establishing knowledge and understanding. Even when you can't get exact and precise knowledge about a topic, you can often find interesting things to learn (for example, why you can't get that knowledge). So, rational inquiry (critical thinking, doing philosophy) may be interesting and fruitful even when we don't get straightforward answers. Once we raise a philosophical issue, a question or puzzle, 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. For example, if my question is 'when is it morally permissible to lie?', I look at cases where it seems OK and cases where it just seems wrong. I look at the **reasons** why they seem that way.

At this point, I am making **arguments**. These are my sets of reasons for a stance. Some arguments give us better reasons or accepting their **conclusions** (what is being argued for) than others. Once we have an argument, we want to evaluate the reasoning it offers. So, if you want to know what philosophers do, this is a pretty good answer: philosophers formulate and evaluate arguments. We look at the reasons for some stance and figure out whether those reasons actually hold water, so to speak. Your introduction to philosophy should be as much a training in how to do philosophy as it is a chance to become acquainted with the views of various philosophers. To that end, you should carefully study the sections below on arguments. I personally approach teaching philosophy from the perspective that there's still a lot going on, I show you the relevance of philosophy to questions which we are puzzled with today (like Artificial Intelligence, the Abortion Debate, the Existence of God, and so on). Through this, you will become accuanted with philosophers, both historical and contemporary, but this is more of a byproduct than a goal.

Once we have some philosophical position to think about, we want to ask what arguments can be made for or against



it (**formulate**). We then want to examine the quality of the arguments (**evaluate**). Evaluating flawed arguments, figuring out why they are wrong, often leads us to other, stronger, arguments and the process of formulating, clarifying, and evaluating arguments continues.

This circular method of question and answer (make an argument, evaluate it, make a stronger one, and repeat) is known as a **dialectic**. A dialectic looks a lot like debate, but a big difference is in the 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.

A dialectic, on the other hand, is aimed at inquiry. The goal is to learn something new about the issue under discussion. Unlike debate, in a 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. For this module, the reading is one of the more famous of Plato's dialogues, namely *The Apology*. 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.

Doing philosophy boils down to thinking really hard, but in a structured and organized way. The thinking must be critical and comprehensive.

The point of arguments is to get at the **truth**. But, in the kind of world we live in today, it is worth taking a look at what philosophers mean by 'truth', before we get too much into arguments.

### Part 1.1.1: Truth

Since both science and philosophy are mainly concerned with getting at knowledge and understanding about the world (though the kind of knowledge may be argued to be different), it is natural to think that both are after the truth about things. There are some interesting and some confused challenges to the idea that philosophy and science are truth oriented. But for now let's assume that rational inquiry is truth oriented and address a couple of questions about truth. Let's focus on just these two:

1. What is it for a claim to be true?
2. How do we determine that a claim is true?

It's important to keep these two questions separate. Questions about how we know whether something is true (like the second question) are epistemic questions. Epistemic questions, which we will return to in a later module, are questions about knowledge, beliefs, reasons, or, to a certain extent, faith. But the question of what it is for something to be true (the first question) is not an epistemic issue. The truth of a claim is quite independent of how or whether we know it to be true. Questions about whether or not something is true are, more or less metaphysical questions.

For example, consider these two claims:

There's extraterrestrial life.

There's no extraterrestrial life.

I assume we don't know which of one of these is true, but surely one of them is, for any sensible claim that you make, it's either true or false. Whichever of these claims is true, its being true doesn't depend in any way on whether or how we know it to be true. So, there is a correct answer to the question "is there extraterrestrial life?" but as it sits, we don't know what that answer is. There are many truths that will never be

known or believed by anyone, and appreciating this is enough to see that the truth of a claim is not relative to belief, knowledge, proof, or any other epistemic notion.<sup>5</sup> Similarly, unless you hold certain philosophic positions about certain things (we will see this in the Relativism/Meta-Ethics Module), just because you believe something, that does not make it true. How much weight you give a belief also doesn't change how true it is. Unlike somethings which you may encounter, like knowledge, certainty, belief, and so on, truth does not come in degrees (and this is regardless of the stances you take).

But then what is it for a claim to be true? The ordinary everyday notion of truth would have it that a claim is true if the world is the way the claim says it is. And this is pretty much all we are after for this class. When we make a claim, we represent some part of the world as being a certain way. If how my claim represents the world fits with the way the world is, then my claim is true. Truth, then, is correspondence, or good fit, between what we assert and the way things are. There are other accounts of truth out there, which do better or worse jobs at various things (for example, this account can't make "Harry Potter is a wizard",<sup>6</sup> "Stepan Arkadyevich Oblonsky's wife is mad at him",<sup>7</sup> or "Sherlock Holmes is a detective"<sup>8</sup> true, because those beings don't exist), but this works for our purposes.

---

<sup>5</sup>There are certain facts within mathematics which are not knowable, this could be because of the physical limitations of the computations or because of the nature of the question which the fact is an answer to. For example, there is a first digit to Graham's number, but it is so large that it's physically impossible to determine.

<sup>6</sup>J. K. Rowling, *Harry Potter And the Sorcerer's Stone* (Arthur A. Levine Books, 1998).

<sup>7</sup>Leo Tolstoy, *Anna Karenina* (Oxford UP, 1980).

<sup>8</sup>Arthur Conan Doyle, *The Complete Sherlock Holmes* (Doubleday & Co., 1930).

### Part 1.1.2: Truth and Meaning

A potential confusion about truth comes from confusing a sentence with the meaning behind it. This is where we get into a very basic introduction into Philosophy of Language. This area of philosophy deals with things such as meaning, reference, and so forth (Philosophy of Language is what I mostly work in, it deals with semantics while Linguistics deals with syntax, but the fields overlap quite a bit). Even when we aren't trying very hard, we can use words and sentences in a ton of different ways and there seems to be some vagueness or ambiguity in natural languages. For example, I can easily make examples where sentences might have two or more different interpretations, like double or triple entendres.

3. GOP grills the IRS chief over lost emails.
4. If your dog poops, you must put it in the trash can.
5. A woman gives birth in the UK every 48 seconds.
6. People wanted for pickling and canning.

All four of the above examples have at least 2 different interpretations and, in each case, one seems true while the other seems false. In Example 3, it could be interpreted to mean that the GOP are having an IRS chief BBQ over a server or it could mean that they are harshly questioning the chief. In Example 4, one way to understand this is that it's telling you to put your dog in the garbage can, while another is telling you to put the poop therein. In Example 5, we could say that there's a single woman who is having a baby every 48 seconds (she must be tired) or that some woman, not necessarily the same woman, is having a child in that time. In Example 6, it could be that we have a cannibal looking to store their meats for the winter using canning and fermenting techniques or it could be that a food fermentation business is looking to hire on some more people in those departments. Given these examples, and how often we misunderstand each other on a daily basis, it might be tempting, therefore, to think that the truth of a sentence

must be relative to its interpretation. In an even more robust example, imagine the following:

Suppose that we all collectively switched the sorts of things the words 'dog' and 'cat' pick-out. So, the word 'dog' now is used to point out the meowing critters and the word 'cat' refers to the barking ones. In this case, it would seem that the sentence 'dogs are canines' would be false and 'dogs are feline' would be true. If we, again, flipped the meanings of the words 'feline' and 'canine', we would get that 'dogs are canines' as true, but for a totally different reason than it was originally.

But does this make truth open for interpretation? Well, no, but in a sense, yes. When we look at a language, from one perspective, we notice that things like words and sentences are nothing more than characters on a screen/page, random collections of sounds, or certain structured gestures (in the case of sign languages). Those things, on their own, don't have meaning. There must be something extra, beyond the sound and signs, which has the meaning. Philosophers call the meaning behind a sentence a proposition. A proposition, itself, is not a sentence or a word, rather it's the meaning behind the sentences.

| Sentence             | Language | Proposition            |
|----------------------|----------|------------------------|
| Snow is white        | English  | that snow is white     |
| Schnee ist weiss     | German   | that snow is white     |
| Nix alba est         | Latin    | that snow is white     |
| La neige est blanche | French   | that the snow is white |

All of the above examples are different sentences, made clear because they are in different languages, but they all express the same proposition. The truth of a sentence is relative to

the truth of the proposition attached to it, propositions are the things which are true or false, and then the truth of the proposition is determined by its correspondence with reality. Translators, at least the good ones, often take the sentence in one language, figure out the proposition connected to it, and then express the same proposition in the necessary language.<sup>9</sup>

So the proposition expressed by a sentence is not itself a linguistic thing. Propositions themselves don't have meaning, but rather they are the meaning behind a sentence. For a bit of language to be open to interpretation is for us to be able to attach different propositions (meanings) to it. But the meanings themselves are not open to further interpretation. And it is the proposition, what is meant by the sentence, which makes the statements, sentences, true or false. So, when I speak of arguments consisting of premises (the supporting evidence and their arrangement in the argument), I am talking about the core meaning behind the sentence, not the sentence itself. If we misinterpret the sentence, then we haven't yet gotten on to the claim being made and hence probably don't fully understand the argument. Getting clear on just what an argument says is critical to the dialectical process.

## Simplified Breakdown

Even if you are exceptionally bright, you probably found the last couple paragraphs rather challenging. That's OK. You might work through them again more carefully and come back to it in a day or two if it's still a struggle. The path to becoming a better critical thinker is more like mountain climbing than a walk in the park, but with this crucial difference: no bones get

---

<sup>9</sup>Sometimes there can be cases where a language has the ability to express a proposition which is not possible to express in another without adding something to that language or giving information beyond what was in the original statement. For example, take Cicero's *Epistulae ad Familiares* 9.22, the examples Cicero gives in the letter of profanity and why they make sense are not translatable without explaining some aspect of Latin phonology and semantics.

broken when you fall off an intellectual cliff. So you are always free to try to scale it again. We can sum up the key points of the last few paragraphs as follows:

We use sentences, bits of language, to express propositions.

The proposition, what is meant by the sentence, represents the world as being some way.

The proposition is true when it represents the world in a way that corresponds to how the world is.

Truth, understood as correspondence between a claim (a proposition) and the way the world is, is not relative to meaning, knowledge, belief, or opinion.

Hopefully we now have a better grip on what it is for a claim to be true. A claim is true just when it represents things as they are. As is frequently the case in philosophy, the real work here was just getting clear on the issue. Once we clearly appreciate the question at hand, the answer seems pretty obvious. So now we can set aside the issue of what truth is and turn to the rather different issue of how to determine what's true.

## **The connection**

Arguments are how philosophers, and scientists, and the rest of us get at the truth. We are using them to structure our reasons and prove that the conclusion is true.

## **Part 1.2: Arguments**

The common-sense, everyday, way to tell whether a claim is true or false is to look at the reasons for or against it. Sometimes our observations give us good reasons. For example, I have a good reason for thinking my bicycle has a flat tire when I see the tire sagging on the rim or hear air hissing out of the tube. But often the business of identifying and evaluating reasons is a bit more involved. Logic is the business of identifying

and evaluating reasons. You do this all the time. You give yourself reasons to choose one shirt over another for a job interview, you reason your way through making a choice about which classes to take, you believe certain things based on evidence, and you give reasons for why something isn't your fault (when you make excuses). In all of these cases and more, you are making **arguments**.

This is very different from how you might use the term 'argument'. In everyday language, we sometimes use the word 'argument' to talk about belligerent shouting matches. If you and a friend have an argument in this sense, things are not going well between the two of you. Logic is not concerned with such teeth-gnashing and hair-pulling. They are not arguments, in our sense; they are just disagreements. For a humorous example of how arguments and disagreements differ, take a look at this funny exchange from Monty Python:<sup>10</sup>

---

<sup>10</sup> "Argument Clinic," writer John Cleese, 1972, BBC1.



Man: Is this the right room for an argument?

Other Man:(John Cleese)

I've told you once.

Man: No you haven't.

Other Man: Yes I have.

M: When?

O: Just now.

M: No you didn't!

O: Yes I did!

M: You didn't! ...

O: Oh I'm sorry, is this a five minute argument, or the full half hour?

M: Ah! (taking out his wallet and paying) Just the five minutes.

O: Just the five minutes. Thank you.

O: Anyway, I did.

M: You most certainly did not!

O: Now let's get one thing quite clear: I most definitely told you!

M: Oh no you didn't!

O: Oh yes I did! ...

M: Oh look, this isn't an argument!

(pause)

O: Yes it is!

M: No it isn't!

(pause)

M: It's just contradiction!

O: No it isn't!

M: It IS!

O: It is NOT! ...

M: (exasperated) Oh, this is futile!!

(pause)

O: No it isn't!

M: Yes it is!

(pause)

M: I came here for a good argument!

O: AH, no you didn't, you came here for an argument!

M: An argument isn't just contradiction.

O: Well! it CAN be!

M: No it can't!

M: *An argument is a connected series of statements intended to establish a proposition.*

O: No it isn't!

M: Yes it is! 'tisn't just contradiction.

O: Look, if I *argue* with you, I must take up a contrary position!

M: Yes but it isn't just saying 'no it isn't'.

O: Yes it is! ...

M: No it ISN'T! *Argument is an intellectual process.* Contradiction is just the automatic gainsaying of anything the other person says.

O: It is NOT!

M: It is! ...

An argument is a reason for taking something to be true. **Arguments** are made out of two or more claims, one of which is a **conclusion**. The conclusion is the claim the argument purports to give a reason for believing. The other claims are the **premises**. The premises of an argument taken together are offered as a reason for believing its conclusion. In the above exchange, the person looking for an argument claims that "An argument is a collected series of statements to establish a definite proposition". This is a close approximation to how philosophers use the term, but a better, more exact definition is:

An argument is a collected series of propositions intended to establish others in the series.

The propositions (statements, in this case) which support others are our premises. The one being supported is the conclusion. That 'intent' bit is a heavy hitter as we will soon see.

Some arguments provide better reasons for believing their conclusions than others. In case you have any doubt about that, consider the following examples:

| Example 1.:                                    | Example 2.:                               |
|------------------------------------------------|-------------------------------------------|
| Sam is a line cook.                            | Sam is a line cook.                       |
| Line cooks generally have good kitchen skills. | Line cooks generally aren't paid well.    |
| Therefore, Same can probably cook well         | Therefore, Sam is probably a billionaire. |

| Example 3.:                               | Example 4.:                               |
|-------------------------------------------|-------------------------------------------|
| Boston is in Massachusetts.               | Boston is in California.                  |
| Massachusetts is east of the Rockies.     | California is west of the Rockies.        |
| Therefore, Boston is east of the Rockies. | Therefore, Boston is west of the Rockies. |

The premises in Example 1 provide pretty good support for thinking Sam can cook well. That is, assuming the premises in the first argument are true, we have a good reason to think that its conclusion is true (at this stage, we are not interesting in whether the premises are actually true, only the structure). Whether or not the argument is any good depends on how well they establish the conclusion, relative to the intent. So, we can say that the reasoning in Example 1 is pretty good (at least). The premises in Example 2 give us no reason to think Sam is a millionaire, let alone a billionaire. So whether or not the premises of an argument support its conclusion is a key issue.

Looking at Examples 3 and 4, we see some similarities. The intent is different (in the first two, the goal was to get something with likelihood, in these, the intent is to get something with certainty), but we seem not to like the second (Example 4) and the first seems OK (Example 3). The main problem with Example 4 is very different than the problem with 2. And, in fact, the issue is an entirely different animal. With Example 2, the issue was with the structure, here, the issue is with the truth. Notice, the structure of the arguments, Examples 3 and 4, are exactly the same, there's no difference in how they are being reasoned. If its premises were true, then we would have a good reason to think the conclusion is true (the best reason, 100% certainty reasoning). That is, the premises do support the conclusion. But the first premise of the second argument just isn't true. Boston is not in California. So the latter pair of arguments suggests another key issue for evaluating arguments.

Good arguments have premises which support the conclusion, the best arguments have true premises.

That is pretty much it. The best arguments are those which have true premises that, when taken together, support its conclusion. So, evaluating an argument involves just these two essential steps:

1. Determine whether or not the premises are true.
2. Determine whether or not the premises support the conclusion (that is, whether we have grounds to think the conclusion is true if all of the premises are true).

Often, figuring out whether the premises of an argument are true involves looking at further arguments for those premises individually. An argument might be the last link in a long chain of reasoning. In this case, the quality of the argument depends on the whole chain. Really high-quality philosophy papers (I don't expect these in this class) often involve an argument and further arguments for each of the premises. And since arguments can have multiple premises, each of which might be supported by further arguments, evaluating one argument might be more involved yet, since its conclusion is really supported by a rich network of reasoning, not just one link and then another. While the potential for complication should be clear, the basic idea should be pretty familiar. Think of the little kid who would constantly ask their parent "why?" after being an explanation. Even at a young age we understood that the reasons for believing one thing can depend on the reasons for believing a great many other things.

However involved the network of reasons supporting a given conclusion might be, it seems that there must be some starting points. That is, it seems there must be some reasons for believing things that don't themselves need to be justified in terms of further reasons. There needs to be some sort of bedrock, ground level, foundation at the bottom. Otherwise the network of supporting reasons would go on without end, and the kid

would happily ask 'why' until the end of time. The problem here is about how do we tell where the ultimate foundations of knowledge and justified belief. This is a big epistemological issue and we will return to it later in the course. For now, let's consider one potential answer we are already familiar with. In the sciences, our complex chains of reasoning seem to proceed from the evidence of the senses.

We think that evidence provides the foundation for our edifice of scientific knowledge. Sounds great for science, but where does this leave philosophy? Does philosophy entirely lack evidence on which its reasoning can be based? Philosophy does have a kind of evidence to work from and that evidence is provided by philosophical problems. When we encounter a problem in philosophy this often tells us that the principles and assumptions that generate that problem can't all be correct. This might seem like just a subtle clue that leaves us far from solving the big mysteries. But clues are evidence just the same. Sensory evidence by itself doesn't tell us as much about the nature of the world as we'd like to suppose. Scientific evidence provides clues, but there remains a good deal of problem solving to do in science as well as in philosophy.

So we can assess the truth or falsity of the premises of an argument by examining evidence or by evaluating further argument in support of the premises. Now we will turn to the other step in evaluating arguments and consider the ways in which premises can support or fail to support their conclusions. The question of support is distinct from the question of whether the premises are true. When we ask whether the premises support the conclusions we are asking whether we'd have grounds for accepting the conclusion assuming the premises are true. In answering this question we will want to apply one of two standards of support: deductive validity or inductive strength.

Part 1.2.1: Deductive Validity

There are two kinds of arguments which we deal with, one is **deductive arguments** and the other is inductive arguments, here we are going to be concerned with deduction. When you are dealing with this kind of argument, the standard for the 'goodness' of the argument is validity. An argument counts as deductive whenever it is aiming at this standard of support. Deductive validity is the strictest standard of support we can uphold. It is also the one which is used in the vast majority of philosophy (with the exception of some, very contemporary, fields).<sup>11</sup> In a deductively **valid** argument, the truth of the premises guarantees the truth of the conclusion. This standard is not concerned with whether the premises are actually true, that is a different standard. Basically, if you assume that the premises are true, the conclusion must be true. Here are two equivalent definitions of deductive validity:

|    |                                                                                                                       |
|----|-----------------------------------------------------------------------------------------------------------------------|
| D  | A valid argument is an argument where if its premises are true, then its conclusion must be true.                     |
| D* | A valid argument is an argument where it is not possible for all of its premises to be true and its conclusion false. |

That (D\*) standard is the more formal way, and exact way, of claiming (D). Here are a few examples of deductively valid arguments:

---

<sup>11</sup>This would be Experimental Philosophy, which involves testing philosophic intuitions. This can be useful when you are trying to get a common person, average Joe view.

| Example 5.                                    | Example 6.                        | Example 7.                             |
|-----------------------------------------------|-----------------------------------|----------------------------------------|
| If Socrates is a man, then Socrates is mortal | All primates are mammals          | If it's wet outside, then it's raining |
| Socrates is a man                             | All humans are primates           | It's not raining                       |
| Therefore, Socrates is mortal                 | Therefore, all humans are mammals | Therefore, it's not wet outside.       |

If you think about these two examples for a moment, it should be clear that there is no possible way for the premises to all be true and the conclusion false. The truth of the conclusion is guaranteed by the truth of the premises. 7, on the other hand, should have given you pause. Yes, that argument is valid, but it's not **sound**. Soundness involves whether or not the premises are in fact true, validity concerns the structure of the argument. If an argument is sound, then it's valid, but not the other way around. In contrast, the following arguments are not valid:

| Example 8.                                                    | Example 9.                                | Example 10.                            |
|---------------------------------------------------------------|-------------------------------------------|----------------------------------------|
| If Socrates the Cat is a man, then Socrates the Cat is mortal | Billy or Sally stole cookies from the jar | If it's wet outside, then it's raining |
| Socrates the Cat is mortal                                    | Billy stole cookies                       | It's not wet outside                   |
| Therefore, Socrates the Cat is man                            | Therefore, Sally didn't steal cookies     | Therefore, it's not raining            |

These examples, again, are not valid, they have a flaw in their reasoning, in arguing this way, you will have an error which will lead you astray. To see why, it might require a bit of imagination, but it's a reasonably simple test. Imagine a case where the premises are true and the conclusion is false. If you can't do it, then the argument is valid, if you can, then there's a

flaw in the reasoning. Think of this test as like a trial by worst-case scenario. For Example 8, Socrates the Cat is obviously mortal, and *were* the cat a man, he still would be mortal, but that doesn't mean that a cat is a man. For Example 11., it's perfectly possible that both Billy and Sally were naughty, bad, kids and stole some cookies from the jar, just because one did it doesn't mean that the other didn't.<sup>12</sup> For example 10, this might be easier to imagine for people who have spent time in the desert: In those sorts of regions, there's a phenomenon which I know as sun-showers. This is where the sun is still shining, the ground is perfectly dry, but there's rain coming down from the sky. So, it's perfectly possible for it to be raining and yet not be wet outside.

Deductive validity is the gold standard for an argument.  
Soundness is the platinum standard.

The deductive arguments we've looked at here are pretty intuitive. We only need to think about whether the conclusion could be false even if the premises were true. But most deductive arguments are not so obvious. These example arguments only use one logical rule and only two supporting premises; most of the ones which are really powerful use several logical rules/forms in conjunction with each other and several more lines of supporting evidence. Logic is the science of deductive validity. For a class on this subject, check out PHIL&120, which is the barebones, no fluff, mathematics of Philosophy. Philosophy has made some historic advances in logic over the past few centuries, with great advancements happening in the last

---

<sup>12</sup>Some languages, like Latin, have two different words to express the different kinds of disjunctions (or-statements). There are inclusive disjunctions, which allow for both to be true and there are exclusive disjunctions, which only allow for only one of them to be true. English uses both but does not have a built in way to tell the difference between them consistently. Context is your best bet for this, but also, sometimes 'either... or...' is used for exclusive and just '... or...' is used for inclusive.



century. If you want to see some of the practical side of their good work, what it's actually been used for, just look at your computer. The background coding for that machine, I know because I worked with it for a time and I know the history behind it, is all based on the logical structures which philosophy has used forever.

### Part 1.2.2: Inductive Strength

—n the section, we looked at deductive arguments and their standards are validity and soundness, but those aren't the only kinds of arguments which are used, other fields use a different kind of argumentation, **inductive arguments**. The standard for the goodness here is strength, not validity. Like with deductive arguments, an argument counts as inductive when it's shooting for this kind of support for the conclusion. Inductive strength is a weaker standard of support we can shoot for. That being said, it's also the standard which is found in science.<sup>13</sup> An inductively strong argument is one where the premises make the conclusion more likely, give probabilistic support. Strength, again, is not concerned with the truth of the premises, that's a different standard still. Here are some examples of inductive arguments:

---

<sup>13</sup>As well as in Experimental Philosophy.

| Example 1:                            | Example 2:                                                                        | Example 3:                                    |
|---------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------|
| Sam is a line cook                    | Most things cancerous to mice are cancerous to humans                             | Oscar was born in North America               |
| Line cooks generally can cook well    | Various chemicals in tobacco products are cancerous to mice                       | Oscar was not born in Mexico                  |
| Therefore, Sam can probably cook well | Therefore, various chemicals in tobacco products are probably cancerous to humans | Therefore, Oscar was probably born in the USA |

Examples 1 and 2 seem like pretty decent arguments. The premises do support the conclusion (we are looking at inductive strength here). But, none of the above arguments are, in fact, valid. It's perfectly possible for the premises to be true and the conclusion false. Sam could be a brand new cook hired because he's the manager's son who has never cooked in his life. The chemicals in tobacco products could be cancerous to mice but not humans.<sup>14</sup> Many arguments give us good reasons for accepting their conclusions even if their premises being true fails to completely guarantee the truth of the conclusion. The intent behind all of these arguments is different. The point of these arguments is not to guarantee the conclusion, but rather to make them more likely to be true. We judge how good such an argument is according to how strong it is. Unlike validity,

---

<sup>14</sup>In this case, we can only look at the information provided in the premises, same with validity, so if you disagree with this sentence, you would be correct, but that's bringing information not given as support in the argument.

strength is a bit more wishy-washy, it comes in degrees, which is why I use the terms 'probable' and 'improbable' in the below definitions:

- |    |                                                                                                                                        |
|----|----------------------------------------------------------------------------------------------------------------------------------------|
| I  | An inductively strong argument is an argument in which if its premises are true, its conclusion is probably true.                      |
| I* | An inductively strong argument is an argument in which it is improbable that its conclusion is false given that its premises are true. |

As in the case of validity, when we say that an argument is strong, we are only claiming that if the premises are true then the conclusion is likely to be true. Corresponding to the notion of deductive soundness, an inductive argument that is both strong and has true premises is called a cogent inductive argument. Unlike the case with deductively sound arguments, it is possible for an inductively cogent argument to have true premises and a false conclusion. When we are asking about the validity of an argument, we are asking whether it's possible for the premises to be true and the conclusion false. When we are asking about the strength of an argument, we are asking about the probability of the conclusion being false if we assume that the premises are true. Possibility does not come in degrees, it's either possible or impossible. Probability does come in degrees. In the simplest case, inductive reasoning involves inferring that something is generally the case from a pattern observed in a limited number of cases.<sup>15</sup>

Suppose we conducted a poll of 1000 Seattle voters. The results showed that 600 of them claimed to be Democrats. We could inductively infer that 60% of the voters in Seattle are Democrats. The results of the poll give a pretty good reason to think that around 60% of the voters in Seattle are Democrats. But the results of the poll don't guarantee this conclusion. It is possible that only 50% of the voters in Seattle are Democrats

---

<sup>15</sup>One way to think about this is that deduction goes from general to particular and induction goes from particular to general.

and Democrats were, just by luck, over represented in the 1000 cases we considered, but it may not be probable.

There are a few factors which tell us how strong an inductive argument is. One is how much evidence we have looked at before inductively generalizing. Our inductive argument above would be stronger if we drew our conclusion from a poll of 100,000 Seattle voters, for instance. And it would be much weaker if we had only polled 100. Similarly, if we were trying to figure out the political stances of all of Washington but solely looked at Enumclaw, we would get a radically different view about our standings than if we had looked at Seattle and Enumclaw.

Also, the strength of an inductive argument depends on how much of the evidence represents the amount in reality. So our inductive argument will be stronger if we randomly select our 1,000 voters from the Seattle phone book than if they are selected from the Ballard phone book (Ballard being a notably liberal neighborhood within Seattle).

So far, we've only discussed inductive generalization, where we identify a pattern in a limited number of cases and draw a more general conclusion about a broader class of cases. Inductive argument comes in other varieties as well. In the example we started with about Sam the line cook, we inductively inferred a prediction about Sam based on a known pattern in a broader class of cases. Argument from analogy is another variety of inductive reasoning that can be quite strong. The strength here is in how many commonalities the two cases have, we will see an argument like this in the section concerning the existence of God.

## **Part 1.3: Common Argument Structures In Philosophy**

Since Philosophy relies on arguments to get at the facts about the world, like science, but we go with the gold-standard for ar-

guments, validity, often arguments will come in certain forms, structures. These forms can be just on their own, or in combinations with others. If you structure your arguments in these ways or you make arguments out of combinations of them, you are guaranteed to have a valid argument, soundness is a different story.

### Part 1.3.1: Modus Ponens

This is another easy and intuitive argument structure. It follows a very similar model to cause-and-effect. If I knock over the glass, it will break; I knocked over the glass; so it broke. Here, I have an if-then statement, which can be phrased in several different ways, I affirm the antecedent (the if-part), and thereby get the consequent (the then-part). But, that doesn't mean it's fool-proof. For example, there's a very common logical fallacy called "affirming the consequent", this is where you have an if-then statement, you affirm the then-part, and thereby think you get the antecedent (the if-part). This is not valid, it's not a good structure. For example, take this quick argument (which is Modus Ponens done properly):

If my car won't start, then there's something wrong with the battery terminals. My car won't start. So, there must be something wrong with the battery terminals.

Using cause-and-effect again, there can be many different possible causes for some event. One of them definitely caused the event, but you can't say with certainty which one caused it just because it happened. For example, if an apple is too heavy, it will fall off a tree. Yes, when an apple is too heavy it will fall from a tree, but loads of other things can cause it to fall, so you can't say that it's weight caused it to fall, it could have been the wind.

### Part 1.3.2: Modus Tollens

Modus Tollens is like Modus Ponens in reverse and negated. In this case, you have an if-then statement, you have that the then-part is false, and so you get that the if-part must be false. This makes sense in the case of cause-and-effect too. You know that one thing would cause another and you know that that other thing didn't happen, so you know that the first thing didn't happen. For example, if I sleep in, I will be late, I wasn't late, so I must not have slept in.

Again, this is a pretty simple idea, but it's easily misused. The fallacy here is called 'denying the antecedent'. This, again, is not good reasoning; you have an if-then statement, you have that the if-part is false, so you think that the then-part must be false. But, just because something wasn't caused by one thing, it doesn't mean that it wasn't caused by another. For example, if God exists, then humans exist; God doesn't exist; therefore, man doesn't exist. Poof, all atheists just disappeared, right? No, they didn't. This is not a valid argument.

Modus Tollens, properly, we will mostly see in this class, as it's the easiest for the sort of teaching style which I employ. For example, take the following:

Ethical Egoism is the stance that the morality of an action is determined by how much it benefits the doer personally. By this theory, an action is moral when it benefits me and immoral when it does not. So, if Ethical Egoism is correct, then donating to charity, which will in no way benefit me, is always morally wrong. But, it seems obvious that the most moral actions we can do are selfless (in this case, donating when it won't benefit me). So, Ethical Egoism is incorrect.

### Part 1.3.3: Disjunctive Syllogism

This is by far the most simple argument structure used in Logic and it's probably the hardest to misuse (think that you have a valid argument, but don't). This does not mean that I have seen bad formulations of this which I will show, but first, here is a story:

Years ago, while I was living in Arizona, I had some car trouble. Periodically, my car just would not start, or it spontaneously let me start it after a few attempts, seemingly randomly. I eventually called my local mechanic and had them take it to a shop. The mechanic, maybe knowing I was a philosophy professor, said the following (and this is a real quote): "Either it's your ECM or it's your fuse-box. We tested and it's not your fuse-box. So, the issue must be your ECM."

Basically, a disjunctive syllogism takes 2 possible options, has that one of them is false, and thereby gets the other. This can be used in various different ways, for example, you could have 3 possible options, show that one of them is false and thereby get that it must be one of the other 2. This is a sort of process of elimination sort of argument structure. This is a very simple, easy, argument structure, but it's still possible to structure it poorly. For example, I have seen people take two options, show that one of them is true and thereby claim that the other is false. This is not a good structure. Disjunctions, in English, without context, don't allow for this sort of move. It is possible for both options to be true. There are some cases where this works, but to play it safe, and stay on the windy side of validity, assume that both can be correct.

### Part 1.3.4: Hypothetical Syllogism

This is the next, basic, argument structure which I will cover with you here. This argument structure does not require you to have anything other than conditionals, you don't need any proven facts. But, at the end, you don't get any proven facts out either, all you get are conditionals. We often use this sort of reasoning when we are thinking about what will happen next if we go through some possible scenario. For example, take the following:

If you give a mouse a cookie, he will ask for a glass of milk. If he asks for a glass of milk, then he will ask for a straw. Therefore, if you give a mouse a cookie, then he will ask for a straw.

When the antecedent of one conditional is the same as the consequent of another, you can collapse them into one conditional by taking out the middle man. This is a driving force for the slippery slope fallacy, but in that case, the fallacy is not in the reasoning itself, but rather in the truth of the conditionals which it employs. Often in arguments, this is used to shorten the work of Modus Ponens or Modus Tollens (below), but it can be used all on its own to get a point across.

### Combining the Structures

Arguments aren't always merely 3 lines long, sometimes they are far longer and more involved. For example, take the following perfectly valid argument:

If I eat a ton of food this Thanksgiving, then I will get yelled at by my doctor. If I get yelled at by my doctor, then I will need to work out. I did eat a ton of food this Thanksgiving. Therefore, I will need to work out.



In this case, I am using several different structures together to get my conclusion. I am using Hypothetical Syllogism and Modus Ponens all at once to get my answer. Similarly, I can have an argument like this:

If I need to get a math credit, then I will either need to take PHIL&120 or a college level math course. I need to get a math credit and I really don't want to take a college level math course, so I will need to take PHIL&120.

Here, I am using both Modus Ponens and Disjunctive Syllogism to get the conclusion. And yes, PHIL&120 is a math credit, but it can be more difficult than other options. All of these examples come from PHIL&120.

## Part 1.4: Fallacies and Biases

A fallacy is just a mistake in reasoning. Humans are not nearly as rational as we'd like to suppose. In fact we are so prone to certain sorts of mistakes in reasoning that philosophers and logicians refer those mistakes by name. For now I will discuss just one by name but in a little detail. Watch for explanations of other fallacies over the course of the class. For pretty thorough catalogue of logical fallacies, I'll refer to you The Fallacy Files.<sup>16</sup> There is also a section in the appendix of this textbook which lists and explains various informal fallacies. Formal fallacies are cases where the structure of the argument seems fine, but it actually relies on an improper move (outlined when I covered the common structures). Informal fallacies are when improper or incorrect reasons are used in the argument, which are outlined in the module I linked to. Here are some examples of fallacies which are rather egregious and should never be seen in any work, in philosophy or elsewhere:

---

<sup>16</sup>Curtis, Gary. What are the fallacy files? *What are the Fallacy Files?* [www.fallacyfiles.org/whatarff.html](http://www.fallacyfiles.org/whatarff.html).

## Ad Hominem Fallacy

“Ad hominem” is Latin for “against the man.” It is the name for the fallacy of attacking the proponent of a position rather than critically evaluating the reasons offered for the proponent’s position. The reason ad hominem is a fallacy is just that the attack on an individual is simply not relevant to the quality of the reasoning offered by that person. Attacking the person who offers an argument has nothing to do whether or not the premises of the argument are true or support the conclusion. Ad hominem is a particularly rampant and destructive fallacy in our society. What makes it so destructive is that it turns the cooperative social project of inquiry through conversation into polarized verbal combat. This fallacy makes rational communication impossible while it diverts attention from interesting issues that often could be fruitfully investigated.

Here is a classic example of ad hominem: A car salesman argues for the quality of an automobile and the potential buyer discounts the argument with the thought that the person is just trying to earn a commission. There may be good reason to think the salesman is just trying to earn a commission. But even if there is, this is irrelevant to the evaluation of the reasons the salesman is offering. The reasons should be evaluated on their own merits.

Notice, it is easy to describe a situation where it is both true that the salesman is just trying to earn a commission and true that he is making good arguments. Consider a salesman who is not too fond of people and cares little for them except that they earn a commission for him. Otherwise he is scrupulously honest and a person of moral integrity. In order to reconcile himself with the duties of a sales job, he carefully researches his product and only accepts a sales position with the business that sells the very best. He then sincerely delivers good arguments for the quality of his product, makes lots of money, and dresses

well. This salesman must have been a philosophy major. The customer who rejects his argument on the ad hominem grounds that he is just trying to earn a commission misses an opportunity to buy the best. The moral of the story is just that the salesperson's motive is logically independent of the quality of his argument.

## Strawman Fallacy

The strawman fallacy is one which I encounter occasionally in student works but you really will see it commonly in the political sphere (much to our misfortune, regardless of the side you hold). A strawman is like a scarecrow, a mock façade of a person used as a distraction or a trick. The strawman in the fallacy is not a façade of the person making the argument or claim, per se, but rather it is a façade of the claim or argument. Often, when we are explaining a view or defending our own views, we will need to explain the opposing side and give their arguments for their stance. This is so that we will be able to explain why our own is more reasonable or better. Putting your opponent's stance in your own words is not strawmanning, in fact I encourage you to (because then you will need to put yourself in their shoes and potentially find flaws in your own stance). This becomes strawmanning, however, when the argument or stance which you are attributing to your opponent is not their argument or stance. You have left out key details, misrepresented their findings or claims, or imposed your own stance (which they reject) onto them in order to make them seem absurd. Strawmanning another person's stance doesn't only harm your quest for the right answer to a problem (because you have removed the view of a person (who I am assuming is reasonable) from the discussion) but it also harms others who could take what you say seriously. In the case of politics, if the party or group you support presents the opposing side as having absurd or outlandish views which could never work or have some fundamental flaw which anyone could see, often,

rather than thinking that they must have missed something, you will take it as gospel and move on to spread this misinformation. The vast majority of the time, however, the absurd or outlandish view was actually a strawman of the original. The original, if the people behind it are reasonable, would lack the fundamental flaw or have some explanation about how the flaw was handled in connection with other policies.

The way to avoid this fallacy is to paint all arguments for any claim in the best possible light. If there is a subtle flaw in the argument which could easily be overlooked, make the patch yourself (even and especially if you disagree with the conclusion to it). For example, in this class, when we talk about the first cause argument for the existence of God, I present the argument without a 'fatal' flaw which is seen in the ordinary presentation of it because it was easy to see and fix; I did not want to present you with a strawman. If a stance seems absurd, do some digging into the actual stance and see whether the person presenting it to you missed or intentionally left out something. For example, suppose that a study showed that some program, which would remove a service from the private sector and made it public (have the government provide the service rather than a collection of private companies), would cost the tax payers 2.6 trillion USD. They present this as an absurd amount of money and then show how much your taxes would increase by. What they fail to mention, however, is that the current cost for that program on tax payers is closer 3.6 trillion USD, but rather than this money going to the government, it goes to the private companies. Others may strawman the cost by inflating it by ignoring the basic fact that many middlemen would not exist in the proposed system and the lack of them would reduce the cost. Doing a little deeper digging will save you from this fallacy.

## Part 1.5: Critical Thinking

This page roughly finishes up our crash course on logic and critical thinking, but we will be seeing more logical forms and critical thinking structures as the course progresses. Most of it was dedicated to logic, and if you are more interested in that, I strongly recommend that you take PHIL&120. But, for good critical thinking, there are some further standards which are worth noting. These standards follow from the others which we have discussed, but are a wee-bit more practical. These standards are Consistency and Coherency, but there are other aspects which come in when philosophy goes empirical.

### Consistency

The first, basic, bare-bones standard for good critical thinking, and thereby good philosophy, is that the thinking must be consistent. This is more basic than saying that it's logical or that it lines-up with reality. For some train of thought to be consistent, it needs to lack two different things. First, it can't have any contradictions. A contradiction is a case where some proposition must be both true and false (at the same time) in order for your reasoning to work. In everyday life, it is (hopefully) very rare for us to encounter a contradiction either in our own or someone else's thinking, but there are cases where they show up, such as in conspiracy theories, and we only notice them well after the stance was explained. Take, for example, a conspiracy theory which has that both the Earth is the center of the universe and that gravity is just the Earth moving upwards at an ever increasing rate. This might not seem too contradictory on the face of it, but once you start digging into what must be true for both of those statements to be true, you will find that some proposition must be both true and false at the same time. There are other more blunt examples, such as:

| Contradictory                                                              | Non-Contradictory                                                          |
|----------------------------------------------------------------------------|----------------------------------------------------------------------------|
| The non-existent ghosts stole the painting                                 | The ghosts stole the painting                                              |
| Bobby Joe committed the crime in New York (in person) while in LA.         | Bobby Joe made it seem like he was in LA in order to cover up the crime.   |
| No one drives in New York because there is too much traffic. <sup>17</sup> | The traffic in New York is so bad that most people walk.                   |
| All animals are equal, but some are more equal than others. <sup>18</sup>  | Either all animals are equal or they aren't all equal.                     |
| I make my own choices, with my wife's permission.                          | What I ultimately choose should be a joint decision between my wife and I. |

The Non-Contradictory column is there to show you how, most of the time, there is an easy way to resolve the contradiction, but there are other cases, which we will see in this class, where the contradiction is more deep seeded and can't be so easily resolved (like the contradictions in Moral Relativism). In order to avoid these, make sure that all of the propositions you are using are consistent, that they are always true or false (depending) throughout your thinking.

The second requirement for consistent thinking is that there aren't any equivocations. This is where you use the same statement or word/phrase in two or more parts of your reasoning and your reasoning relies on them meaning the same thing, but they don't mean the same thing in the two different instances. For example, take a look at these arguments:

Nothing is better than God.

A cheese sandwich is better than nothing.

Therefore, a cheese sandwich is better than God.

In this argument, the term 'nothing' is being used in two different ways. In order for us to accept the two lines of this argument, individually, we have different meanings in mind. In the first case, 'nothing' means that something along the lines of 'it is not the case that there exists a being such that that being...' or "no being". So, the proper way to understand this first line is "it is not the case that there exists a being such that that being is better than God." The second line of the argument uses a different sense of the word 'nothing', namely "not having anything", so the correct way to understand this is that it means "a cheese sandwich is better than not having anything." Taking these two together, the argument looks like this:

There isn't anything better than God.

Having a cheese sandwich is better than not having anything.

Therefore, ...

As you should be able to see, the flow of the argument isn't there any more, the equivocation was the glue holding it together. Some other examples can be in quick reasoning like:

André the Giant is so called because of his great size.

André René Roussimoff is so called because that's what his mother named him.

André the Giant is André René Roussimoff.

Therefore, André René Roussimoff is so called because of his great size.

That last line should seem off to you. This is because there is equivocation in this reasoning. In the first two lines we are talking about the names, but in the third we are talking about what the names pick out in the world. The equivocation is in the shift between talking about the word and the person.

## Coherency

The next basic requirement for good critical thinking is that it needs to be coherent. Coherency is not the same as consistency, though there is a fair bit of overlap. Each premise, each proposition used in the reasoning need to relate to each other in a reasonable way. There can't be any strange or unorthodox jumps in the argument. You can think of this like a spider web, each of the strands is connected together and it would not be as strong if some of those strands were removed. For example, take this argument:

If the moon landing was fake, then the Government did so to deceive us.

If the Government did so to deceive us, then it was to make us lose faith in our religion.

The moon landing was fake.

Therefore, the Government faked it in order to make us lose faith in our religion.

So, I am not going to make an argument that the moon landing wasn't fake (we really did go to the moon), rather, assuming that it was fake, it does make sense that it would have been faked to deceive us in some way. The real issue with this argument is in the second line. Does it really make sense that the Government would deceive us to make us lose faith in this way? There is a jump in the reasoning. Much more evidence and premises are required to show a connection between the assumption that the Government deceived us with the moon landing and that the purpose of it had something to do with religion. There are many other closer, more relevant potential reasons which need to be discounted first. For example, that the Government did so in order to reaffirm a sense of exceptionalism.



## Empirical Thinking

There was a time in which most philosophic thought was empirical, but as the sciences diversified, that became less common (though the trend is swinging back that way in philosophy). Empirical thinking is reasoning which relates to or explains the outside world. In the other sciences, you may have heard a distinction between empirical (applied) and theoretical. Theoretical science is the arm-chair, hypothetical models which are thought about and debated and later tested using empirical methods (like experiments). Empirical science is the science which actually does the tests and uses the materials in question. There are two features to reasonable empirical thinking, which the other sciences should take note of and explain from the beginning. These are that the thinking must be adequate and applicable.

Empirical thinking is adequate when all of the cases you are trying to explain are accounted for. There shouldn't be too many exceptions to the account you are giving and those exceptions should be easily accounted for. For example, although this story is apocryphal, when Galileo presented his findings about there being objects which orbit something other than the Earth, one person claimed that it was because of interference in the telescope. So, they tried it in a different location, and they got the same result. Eventually, the objector said something like "my hypothesis works so long as you don't look through a telescope". The hypothesis that all objects orbit the Earth wasn't adequate because there were cases which it could not explain and those exceptions could not be easily accounted for.

Empirical thinking is applicable when there isn't anything in the explanation which doesn't relate back to experience and evidence or data gained from testing in the relevant environment and the explanation/hypothesis is useful (as in, it leads or gives way to further understanding and can be used to make other explanations). Many examples of empirical thoughts failing in this regard can be found when you look at the claims

and tests involving pyramid power, magic crystals, feng shui, various vitamins and supplements seen on Dr. Oz, and even copper/magnetic bracelets claimed to treat arthritis. The explanations for each of these contain claims which either can't be demonstrated with experiments or don't relate to experience. For a more scientific example, take this quote from the famous physicist Richard Feynman:

The next question was — what makes planets go around the sun? At the time of Kepler some people answered this problem by saying that there were angels behind them beating their wings and pushing the planets around an orbit.<sup>a</sup>

---

<sup>a</sup>R. P. Feynman, *In The character of physical law* (The MIT P, 2017).

The hypothesis that the angels are making the planets move isn't applicable because it doesn't relate back to experience and evidence. If, in some strange world, we could see the angels sweating and pushing the planets really hard, then it would be. However, we can't see the angels and if someone were to claim that they are there, they are just invisible, the hypothesis would be even more inapplicable.

## *Plato's Apology*

I do not know, men of Athens, how my accusers affected you; 17  
as for me, I was almost carried away in spite of myself, so per-  
suasively did they speak. And yet, hardly anything of what  
they said is true. Of the many lies they told, one in particular  
surprised me, namely that you should be careful not to be de-  
ceived by an accomplished speaker like me. That they were not  
ashamed to be immediately proved wrong by the facts, when b  
I show myself not to be an accomplished speaker at all, that I  
thought was most shameless on their part—unless indeed they  
call an accomplished speaker the man who speaks the truth.  
If they mean that, I would agree that I am an orator, but not  
after their manner, for indeed, as I say, practically nothing they  
said was true. From me you will hear the whole truth, though c  
not, by Zeus, gentlemen, expressed in embroidered and stylized  
phrases like theirs, but things spoken at random and expressed  
in the first words that come to mind, for I put my trust in the  
justice of what I say, and let none of you expect anything else.  
It would not be fitting at my age, as it might be for a young  
man, to toy with words when I appear before you.

One thing I do ask and beg of you, gentlemen: if you hear  
me making my defence in the same kind of language as I am  
accustomed to use in the market place by the bankers' tables,  
where many of you have heard me, and elsewhere, do not be  
surprised or create a disturbance on that account. The posi-

tion is this: this is my first appearance in a lawcourt, at the age of seventy; I am therefore simply a stranger to the manner of d speaking here. Just as if I were really a stranger, you would certainly excuse me if I spoke in that dialect and manner in which I had been brought up, so too my present request seems a just one, for you to pay no attention to my manner of speech—be it 18 better or worse—but to concentrate your attention on whether what I say is just or not, for the excellence of a judge lies in this, as that of a speaker lies in telling the truth.

It is right for me, gentlemen, to defend myself first against the first lying accusations made against me and my first accusers, and then against the later accusations and the later accusers. There have been many who have accused me to you for many years now, and none of their accusations are true. These b I fear much more than I fear Anytus and his friends, though they too are formidable. These earlier ones, however, are more so, gentlemen; they got hold of most of you from childhood, persuaded you and accused me quite falsely, saying that there is a man called Socrates, a wise man, a student of all things in the sky and below the earth, who makes the worse argument the stronger. Those who spread that rumour, gentlemen, are my dangerous accusers, for their hearers believe that those who c study these things do not even believe in the gods. Moreover, these accusers are numerous, and have been at it a long time; also, they spoke to you at an age when you would most readily believe them, some of you being children and adolescents, and they won their case by default, as there was no defence.

What is most absurd in all this is that one cannot even know or mention their names unless one of them is a writer of comedies. Those who maliciously and slanderously persuaded you—who also, when persuaded themselves then persuaded oth- d ers—all those are most difficult to deal with: one cannot bring one of them into court or refute him; one must simply fight with shadows, as it were, in making one's defence, and cross-examine when no one answers. I want you to realize too that my accusers are of two kinds: those who have accused me re-

cently, and the old ones I mention; and to think that I must first defend myself against the latter, for you have also heard their accusations first, and to a much greater extent than the more recent.

Very well then. I must surely defend myself and attempt to uproot from your minds in so short a time the slander that has resided there so long. I wish this may happen, if it is in any way better for you and me, and that my defence may be successful, but I think this is very difficult and I am fully aware of how difficult it is. Even so, let the matter proceed as the god may wish, but I must obey the law and make my defence.

Let us then take up the case from its beginning. What is the accusation from which arose the slander in which Meletus trusted when he wrote out the charge against me? What did they say when they slandered me? I must, as if they were my actual prosecutors, read the affidavit they would have sworn. It goes something like this: Socrates is guilty of wrongdoing in that he busies himself studying things in the sky and below the earth; he makes the worse into the stronger argument, and he teaches these same things to others. You have seen this yourselves in the comedy of Aristophanes, a Socrates swinging about there, saying he was walking on air and talking a lot of other nonsense about things of which I know nothing at all. I do not speak in contempt of such knowledge, if someone is wise in these things—lest Meletus bring more cases against me—but, gentlemen, I have no part in it, and on this point I call upon the majority of you as witnesses. I think it right that all those of you who have heard me conversing, and many of you have, should tell each other if anyone of you has ever heard me discussing such subjects to any extent at all. From this you will learn that the other things said about me by the majority are of the same kind.

Not one of them is true. And if you have heard from anyone that I undertake to teach people and charge a fee for it, that is not true either. Yet I think it a fine thing to be able to teach people as Gorgias of Leontini does, and Prodicus of Ceos,

and Hippias of Elis. Each of these men can go to any city and persuade the young, who can keep company with anyone of their own fellow-citizens they want without paying, to leave the company of these, to join with themselves, pay them a fee, and be grateful to them besides. Indeed, I learned that there is another wise man from Paros who is visiting us, for I met a man who has spent more money on Sophists than everybody else put together, Callias, the son of Hipponicus. So I asked him—he has two sons—"Callias," I said, "if your sons were colts or calves, we could find and engage a supervisor for them who would make them excel in their proper qualities, some horse breeder or farmer. Now since they are men, whom do you have in mind to supervise them? Who is an expert in this kind of excellence, the human and social kind? I think you must have given thought to this since you have sons. Is there such a person," I asked, "or is there not?" "Certainly there is," he said. "Who is he?" I asked, "What is his name, where is he from? and what is his fee?" "His name, Socrates, is Evenus, he comes from Paros, and his fee is five minas." I thought Evenus a happy man, if he really possesses this art, and teaches for so moderate a fee. Certainly I would pride and preen myself if I had this knowledge, but I do not have it, gentlemen.

One of you might perhaps interrupt me and say: "But Socrates, what is your occupation? From where have these slanders come? For surely if you did not busy yourself with something out of the common, all these rumours and talk would not have arisen unless you did something other than most people. Tell us what it is, that we may not speak inadvisedly about you." Anyone who says that seems to be right, and I will try to show you what has caused this reputation and slander. Listen then. Perhaps some of you will think I am jesting, but be sure that all that I shall say is true. What has caused my reputation is none other than a certain kind of wisdom. What kind of wisdom? Human wisdom, perhaps. It may be that I really possess this, while those whom I mentioned just now are wise with a wisdom more than human; else I cannot explain it, for I

certainly do not possess it, and whoever says I do is lying and speaks to slander me. Do not create a disturbance, gentlemen, even if you think I am boasting, for the story I shall tell does not originate with me, but I will refer you to a trustworthy source. I shall call upon the god at Delphi as witness to the existence and nature of my wisdom, if it be such. You know Chairephon. He was my friend from youth, and the friend of most of you, as he shared your exile and your return. You surely know the kind of man he was, how impulsive in any course of action. He went to Delphi at one time and ventured to ask the oracle—as I say, gentlemen, do not create a disturbance—he asked if any man was wiser than I, and the Pythian replied that no one was wiser. Chairephon is dead, but his brother will testify to you about this. 21

Consider that I tell you this because I would inform you about the origin of the slander. When I heard of this reply I asked myself: "Whatever does the god mean? What is his riddle? I am very conscious that I am not wise at all; what then does he mean by saying that I am the wisest? For surely he does not lie; it is not legitimate for him to do so." For a long time I was at a loss as to his meaning; then I very reluctantly turned to some such investigation as this: I went to one of those reputed wise, thinking that there, if anywhere, I could refute the oracle and say to it: "This man is wiser than I, but you said I was." Then, when I examined this man—there is no need for me to tell you his name, he was one of our public men—my experience was something like this: I thought that he appeared wise to many people and especially to himself, but he was not. I then tried to show him that he thought himself wise, but that he was not. As a result he came to dislike me, and so did many of the bystanders. So I withdrew and thought to myself: "I am wiser than this man; it is likely that neither of us knows anything worthwhile, but he thinks he knows something when he does not, whereas when I do not know, neither do I think I know; so I am likely to be wiser than he to this small extent, that I do not think I know what I do not know." After this I

approached another man, one of those thought to be wiser than he, and I thought the same thing, and so I came to be disliked both by him and by many others.

After that I proceeded systematically. I realized, to my sorrow and alarm, that I was getting unpopular, but I thought that I must attach the greatest importance to the god's oracle, so I must go to all those who had any reputation for knowledge to examine its meaning. And by the dog, gentlemen of the jury—for I must tell you the truth—I experienced something like this: in my investigation in the service of the god I found that those who had the highest reputation were nearly the most deficient, while those who were thought to be inferior were more knowledgeable. I must give you an account of my journeyings as if they were labours I had undertaken to prove the oracle irrefutable. After the politicians, I went to the poets, the writers of tragedies and dithyrambs and the others, intending in their case to catch myself being more ignorant than they. So I took up those poems with which they seemed to have taken most trouble and asked them what they meant, in order that I might at the same time learn something from them. I am ashamed to tell you the truth, gentlemen, but I must. Almost all the bystanders might have explained the poems better than their authors could. I soon realized that poets do not compose their poems with knowledge, but by some inborn talent and by inspiration, like seers and prophets who also say many fine things without any understanding of what they say. The poets seemed to me to have had a similar experience. At the same time I saw that, because of their poetry, they thought themselves very wise men in other respects, which they were not. So there again I withdrew, thinking that I had the same advantage over them as I had over the politicians.

Finally I went to the craftsmen, for I was conscious of knowing practically nothing, and I knew that I would find that they had knowledge of many fine things. In this I was not mistaken; they knew things I did not know, and to that extent they were wiser than I. But, gentlemen of the jury, the good craftsmen



seemed to me to have the same fault as the poets: each of them, because of his success at his craft, thought himself very wise in other most important pursuits, and this error of theirs overshadowed the wisdom they had, so that I asked myself, on behalf of the oracle, whether I should prefer to be as I am, with neither their wisdom nor their ignorance, or to have both. The answer I gave myself and the oracle was that it was to e my advantage to be as I am. As a result of this investigation, gentlemen of the jury, I acquired much unpopularity, of a kind that is hard to deal with and is a heavy burden; many slanders came from these people and a reputation for wisdom, for 23 in each case the bystanders thought that I myself possessed the wisdom that I proved that my interlocutor did not have. What is probable, gentlemen, is that in fact the god is wise and that his oracular response meant that human wisdom is worth little or nothing, and that when he says this man, Socrates, he is using my name as an example, as if he said: "This man among you, mortals, is wisest who, like Socrates, understands that his wisdom is worthless." So even now I continue this investigation b as the god bade me—and I go around seeking out anyone, citizen or stranger, whom I think wise. Then if I do not think he is, I come to the assistance of the god and show him that he is not wise. Because of this occupation, I do not have the leisure to engage in public affairs to any extent, nor indeed to look after my own, but I live in great poverty because of my service to the god.

Furthermore, the young men who follow me around of their own free will, those who have most leisure, the sons of the very rich, take pleasure in hearing people questioned; they themselves often imitate me and try to question others. I think they c find an abundance of men who believe they have some knowledge but know little or nothing. The result is that those whom they question are angry, not with themselves but with me. They say: "That man Socrates is a pestilential fellow who corrupts the young." If one asks them what he does and what he teaches to corrupt them, they are silent, as they do not know, but, so d

as not to appear at a loss, they mention those accusations that are available against all philosophers, about "things in the sky and things below the earth," about "not believing in the gods" and "making the worse the stronger argument;" they would not want to tell the truth, I'm sure, that they have been proved to lay claim to knowledge when they know nothing. These people are ambitious, violent and numerous; they are continually and convincingly talking about me; they have been filling your ears for a long time with vehement slanders against me. From them Meletus attacked me, and Anytus and Lycon, Meletus being vexed on behalf of the poets, Anytus on behalf of the craftsmen e and the politicians, Lycon on behalf of the orators, so that, as I started out by saying, I should be surprised if I could rid you of so much slander in so short a time. That, gentlemen of the jury, is the truth for you. I have hidden or disguised nothing. I 24 know well enough that this very conduct makes me unpopular, and this is proof that what I say is true, that such is the slander against me, and that such are its causes. If you look into this either now or later, this is what you will find. b

Let this suffice as a defence against the charges of my earlier accusers. After this I shall try to defend myself against Meletus, that good and patriotic man, as he says he is, and my later accusers. As these are a different lot of accusers, let us again take up their sworn deposition. It goes something like this: Socrates is guilty of corrupting the young and of not believing in the gods in whom the city believes, but in other new spiritual things. Such is their charge. Let us examine it point by point. c

He says that I am guilty of corrupting the young, but I say that Meletus is guilty of dealing frivolously with serious matters, of irresponsibly bringing people into court, and of professing to be seriously concerned with things about none of which he has ever cared, and I shall try to prove that this is so. Come here and tell me, Meletus. Surely you consider it of the greatest importance that our young men be as good as possible? —Indeed d I do.

Come then, tell the jury who improves them. You obviously

know, in view of your concern. You say you have discovered the one who corrupts them, namely me, and you bring me here and accuse me to the jury. Come, inform the jury and tell them who it is. You see, Meletus, that you are silent and know not what to say. Does this not seem shameful to you and a sufficient proof of what I say, that you have not been concerned with any of this? Tell me, my good sir, who improves our young men? —The laws.

That is not what I am asking, but what person who has knowledge of the laws to begin with?—These jurymen, Socrates.

How do you mean, Meletus? Are these able to educate the young and improve them?—Certainly.

All of them, or some but not others?—All of them.

Very good, by Hera. You mention a great abundance of benefactors. But what about the audience? Do they improve the young or not?—They do, too.

What about the members of Council?—The Councillors, also. But, Meletus, what about the assembly? Do members of the assembly corrupt the young, or do they all improve them?—They improve them.

All the Athenians, it seems, make the young into fine good men, except me, and I alone corrupt them. Is that what you mean?—That is most definitely what I mean.

You condemn me to a great misfortune. Tell me: does this also apply to horses do you think? That all men improve them and one individual corrupts them? Or is quite the contrary true, one individual is able to improve them, or very few, namely the horse breeders, whereas the majority, if they have horses and use them, corrupt them? Is that not the case, Meletus, both with horses and all other animals? Of course it is, whether you and Anytus say so or not. It would be a very happy state of affairs if only one person corrupted our youth, while the others improved them.

You have made it sufficiently obvious, Meletus, that you have never had any concern for our youth; you show your indif-

ference clearly; that you have given no thought to the subjects about which you bring me to trial.

And by Zeus, Meletus, tell us also whether it is better for a man to live among good or wicked fellow-citizens. Answer, my good man, for I am not asking a difficult question. Do not the wicked do some harm to those who are ever closest to them, whereas good people benefit them?—Certainly.

And does the man exist who would rather be harmed than benefited by his associates? Answer, my good sir, for the law orders you to answer. Is there any man who wants to be harmed?—Of course not.

Come now, do you accuse me here of corrupting the young and making them worse deliberately or unwillingly?—Deliberately.

What follows, Meletus? Are you so much wiser at your age than I am at mine that you understand that wicked people always do some harm to their closest neighbors while good people do them good, but I have reached such a pitch of ignorance that I do not realize this, namely that if I make one of my associates wicked I run the risk of being harmed by him so that I do such a great evil deliberately, as you say? I do not believe you, Meletus, and I do not think anyone else will. Either I do not corrupt the young or, if I do, it is unwillingly, and you are lying in either case. Now if I corrupt them unwillingly, the law does not require you to bring people to court for such unwilling wrongdoings, but to get hold of them privately, to instruct them and exhort them; for clearly, if I learn better, I shall cease to do what I am doing unwillingly. You, however, have avoided my company and were unwilling to instruct me, but you bring me here, where the law requires one to bring those who are in need of punishment, not of instruction.

And so, gentlemen of the jury, what I said is clearly true: Meletus has never been at all concerned with these matters. Nonetheless tell us, Meletus, how you say that I corrupt the young; or is it obvious from your deposition that it is by teaching them not to believe in the gods in whom the city believes

but in other new spiritual things? Is this not what you say I teach and so corrupt them? —That is most certainly what I do say.

Then by those very gods about whom we are talking, Meletus, make this clearer to me and to the jury: I cannot be sure whether you mean that I teach the belief that there are some gods—and therefore I myself believe that there are gods and am not altogether an atheist, nor am I guilty of that—not, however, the gods in whom the city believes, but others, and that this is the charge against me, that they are others. Or whether you mean that I do not believe in gods at all, and that this is what I teach to others. —This is what I mean, that you do not believe in gods at all.

You are a strange fellow, Meletus. Why do you say this? Do I not believe, as other men do, that the sun and the moon are gods?—No, by Zeus, jurymen, for he says that the sun is stone, and the moon earth.

My dear Meletus, do you think you are prosecuting Anaxagoras? Are you so contemptuous of the jury and think them so ignorant of letters as not to know that the books of Anaxagoras of Clazomenae are full of those theories, and further, that the young men learn from me what they can buy from time to time for a drachma, at most, in the bookshops, and ridicule Socrates if he pretends that these theories are his own, especially as they are so absurd? Is that, by Zeus, what you think of me, Meletus, that I do not believe that there are any gods? —That is what I say, that you do not believe in the gods at all.

You cannot be believed, Meletus, even, I think, by yourself. The man appears to me, gentlemen of the jury, highly insolent and uncontrolled. He seems to have made this deposition out of insolence, violence and youthful zeal. He is like one who composed a riddle and is trying it out: "Will the wise Socrates realize that I am jesting and contradicting myself, or shall I deceive him and others?" I think he contradicts himself in the affidavit, as if he said: "Socrates is guilty of not believing in 27

gods but believing in gods," and surely that is the part of a jester!

Examine with me, gentlemen, how he appears to contradict himself, and you, Meletus, answer us. Remember, gentlemen, what I asked you when I began, not to create a disturbance if I proceed in my usual manner. b

Does any man, Meletus, believe in human activities who does not believe in humans? Make him answer, and not again and again create a disturbance. Does any man who does not believe in horses believe in horsemen's activities? Or in flute-playing activities but not in flute-players? No, my good sir, no man could. If you are not willing to answer, I will tell you and the jury. Answer the next question, however. Does any man believe in spiritual activities who does not believe in spirits?—No one. c

Thank you for answering, if reluctantly, when the jury made you. Now you say that I believe in spiritual things and teach about them, whether new or old, but at any rate spiritual things according to what you say, and to this you have sworn in your deposition. But if I believe in spiritual things I must quite inevitably believe in spirits. Is that not so? It is indeed. I shall assume that you agree, as you do not answer. Do we not believe spirits to be either gods or the children of gods? Yes or no?—Of course. d

Then since I do believe in spirits, as you admit, if spirits are gods, this is what I mean when I say you speak in riddles and in jest, as you state that I do not believe in gods and then again that I do, since I do believe in spirits. If on the other hand the spirits are children of the gods, bastard children of the gods by nymphs or some other mothers, as they are said to be, what man would believe children of the gods to exist, but not gods? That would be just as absurd as to believe the young of horses and asses, namely mules, to exist, but not to believe in the existence of horses and asses. You must have made this deposition, Meletus, either to test us or because you were at a loss to find any true wrongdoing of which to accuse e

me. There is no way in which you could persuade anyone of even small intelligence that it is possible for one and the same man to believe in spiritual but not also in divine things, and then again for that same man to believe neither in spirits nor in gods nor in heroes.

28

I do not think, gentlemen of the jury, that it requires a prolonged defence to prove that I am not guilty of the charges in Meletus' deposition, but this is sufficient. On the other hand, you know that what I said earlier is true, that I am very unpopular with many people. This will be my undoing, if I am undone, not Meletus or Anytus but the slanders and envy of many people. This has destroyed many other good men and will, I think, continue to do so. There is no danger that it will stop at me. b

Someone might say: 'Are you not ashamed, Socrates, to have followed the kind of occupation that has led to your being now in danger of death?' However, I should be right to reply to him: "You are wrong, sir, if you think that a man who is any good at all should take into account the risk of life or death; he should look to this only in his actions, whether what he does is right or wrong, whether he is acting like a good or a bad man." According to your view, all the heroes who died at Troy were inferior people, especially the son of Thetis who was so contemptuous of danger compared with disgrace. c When he was eager to kill Hector, his goddess mother warned him, as I believe, in some such words as these: "My child, if you avenge the death of your comrade, Patroclus, and you kill Hector, you will die yourself, for your death is to follow immediately after Hector's." Hearing this, he despised death and danger and was much more afraid to live a coward who did not avenge his friends. "Let me die at once," he said, "when once I have given the wrongdoer his deserts, rather than remain d here, a laughing-stock by the curved ships, a burden upon the earth." Do you think he gave thought to death and danger?

This is the truth of the matter, gentlemen of the jury: wherever a man has taken a position that he believes to be best, or

has been placed by his commander, there he must I think remain and face danger, without a thought for death or anything else, rather than disgrace. It would have been a dreadful way to behave, gentlemen of the jury, if, at Potidaea, Amphipolis and Delium, I had, at the risk of death, like anyone else, remained at my post where those you had elected to command had ordered me, and then, when the god ordered me, as I thought and believed, to live "the life of a philosopher, to examine myself and others, I had abandoned my post for fear of death or anything else. That would have been a dreadful thing, and then I might truly have justly been brought here for not believing that there are gods, disobeying the oracle, fearing death, and thinking I was wise when I was not. To fear death, gentlemen, is no other than to think oneself wise when one is not, to think one knows what one does not know. No one knows whether death may not be the greatest of all blessings for a man, yet men fear it as if they knew that it is the greatest of evils. And surely it is the most blameworthy ignorance to believe that one knows what one does not know. It is perhaps on this point and in this respect, gentlemen, that I differ from the majority of men, and if I were to claim that I am wiser than anyone in anything, it would be in this that as I have no adequate knowledge of things in the underworld, so I do not think I have. I do know, however, that it is wicked and shameful to do wrong, to disobey one's superior, be he god or man. I shall never fear or avoid things of which I do not know, whether they may not be good rather than things that I know to be bad. Even if you acquitted me now and did not believe Anytus, who said to you that either I should not have been brought here in the first place, or that now I am here, you cannot avoid executing me, for if I should be acquitted, your sons would practise the teachings of Socrates and all be thoroughly corrupted; if you said to me in this regard: "Socrates, we do not believe Anytus now; we acquit you, but only on condition that you spend no more time on this investigation and do not practise philosophy, and if you are caught doing so you will die;" if, as I say, you



were to acquit me on those terms, I would say to you: "Gentlemen of the jury, I am grateful and I am your friend, but I d  
 will obey the god rather than you, and as long as I draw breath  
 and am able, I shall not cease to practise philosophy, to exhort  
 you and in my usual way to point out to anyone of you whom  
 I happen to meet: Good Sir, you are an Athenian, a citizen of  
 the greatest city with the greatest reputation for both wisdom  
 and power; are you not ashamed of your eagerness to possess  
 as much wealth, reputation and honours as possible, while you  
 do not care for nor give thought to wisdom or truth or the best  
 possible state of your soul?" Then, if one of you disputes this e  
 and says he does care, I shall not let him go at once or leave  
 him, but I shall question him, examine him and test him, and  
 if I do not think he has attained the goodness that he says he  
 has, I shall reproach him because he attaches little importance  
 to the most important things and greater importance to inferior  
 things. I shall treat in this way anyone I happen to meet, young  
 and old, citizen and stranger, and more so the citizens because  
 you are more kindred to me. Be sure that this is what the god 30  
 orders me to do, and I think there is no greater blessing for the  
 city than my service to the god. For I go around doing nothing  
 but persuading both young and old among you not to care for  
 your body or your wealth in preference to or as strongly as for  
 the best possible state of your soul, as I say to you: "Wealth  
 does not bring about excellence, but excellence makes wealth b  
 and everything else good for men, both individually and collec-  
 tively."

Now if by saying this I corrupt the young, this advice must  
 be harmful, but if anyone says that I give different advice, he is  
 talking nonsense. On this point I would say to you, gentlemen  
 of the jury: "Whether you believe Anytus or not, whether you  
 acquit me or not, do so on the understanding that this is my  
 course of action, even if I am to face death many times." Do  
 not create a disturbance, gentlemen, but abide by my request c  
 not to cry out at what I say but to listen, for I think it will be  
 to your advantage to listen, and I am about to say other things

at which you will perhaps cry out. By no means do this. Be sure that if you kill the sort of man I say I am, you will not harm me more than yourselves. Neither Meletus nor Anytus can harm me in any way; he could not harm me, for I do not think it is permitted that a better man be harmed by a worse; certainly he might kill me, or perhaps banish or disfranchise me, which he and maybe others think to be great harm, but I do not think so. I think he is doing himself much greater harm doing what he is doing now, attempting to have a man executed unjustly. Indeed, gentlemen of the jury, I am far from making it defence now on my own behalf, as might be thought, but on yours, to prevent you from wrongdoing by mistreating the god's gift to you by condemning me; for if you kill me you will not easily find another like me. I was attached to this city by the god—though it seems a ridiculous thing to say—as upon a great and noble horse which was somewhat sluggish because of its size and needed to be stirred up by a kind of gadfly. It is to fulfill some such function that I believe the god has placed me in the city. I never cease to rouse each and everyone of you, to persuade and reproach you all day long and everywhere I find myself in your company.

Another such man will not easily come to be among you, gentlemen, and if you believe me you will spare me. You might easily be annoyed with me as people are when they are aroused from a doze, and strike out at me; if convinced by Anytus you could easily kill me, and then you could sleep on for the rest of your days, unless the god, in his care for you, sent you someone else. That I am the kind of person to be a gift of the god to the city you might realize from the fact that it does not seem like human nature for me to have neglected all my own affairs and to have tolerated this neglect now for so many years while I was always concerned with you, approaching each one of you like a father or an elder brother to persuade you to care for virtue (*aretē*). Now if I profited from this by charging a fee for my advice, there would be some sense to it, but you can see for yourselves that, for all their shameless accusations, my

accusers have not been able in their impudence to bring forward a witness to say that I have ever received a fee or ever asked for one. I, on the other hand, have a convincing witness that I speak the truth, my poverty.

It may seem strange that while I go around and give this advice privately and interfere in private affairs, I do not venture to go to the assembly and there advise the city. You have heard me give the reason for this in many places. I have a divine or spiritual sign which Meletus has ridiculed in his deposition. This began when I was a child. It is a voice, and whenever it speaks it turns me away from something I am about to do, but it never encourages me to do anything. This is what has prevented me from taking part in public affairs, and I think it was quite right to prevent me. Be sure, gentlemen of the jury, that if I had long ago attempted to take part in politics, I should have died long ago, and benefited neither you nor myself. Do not be angry with me for speaking the truth; no man will survive who genuinely opposes you or any other crowd and prevents the occurrence of many unjust and illegal happenings in the city. A man who really fights for justice must lead a private, not a public, life if he is to survive for even a short time.

I shall give you great proofs of this, not words but what you esteem, deeds. Listen to what happened to me, that you may know that I will not yield to any man contrary to what is right, for fear of death, even if I should die at once for not yielding. The things I shall tell you are commonplace and smack of the lawcourts, but they are true. I have never held any other office in the city, but I served as a member of the Council, and our tribe Antiochis was presiding at the time when you wanted to try as a body the ten generals who had failed to pick up the survivors of the naval battle. This was illegal, as you all recognized later. I was the only member of the presiding committee to oppose your doing something contrary to the laws, and I voted against it. The orators were ready to prosecute me and take me away; and your shouts were egging them on, but I thought I should run any risk on the side of law and justice

rather than join you, for fear of prison or death, when you were engaged in an unjust course.

This happened when the city was still a democracy. When the oligarchy was established, the Thirty summoned me to the Hall, along with four others, and ordered us to bring Leon from Salamis, that he might be executed. They gave many such orders to many people, in order to implicate as many as possible in their guilt. Then I showed again, not in words but in action, that, if it were not rather vulgar to say so, death is something I couldn't care less about, but that my whole concern is not to do anything unjust or impious. That government, powerful as it was, did not frighten me into any wrongdoing. When we left the Hall, the other four went to Salamis and brought in Leon, but I went home. I might have been put to death for this, had not the government fallen shortly afterwards. There are many who will witness to these events.

Do you think I would have survived all these years if I were engaged in public affairs and, acting as a good man must, came to the help of justice and considered this the most important thing? Far from it, gentlemen of the jury, nor would any other man. Throughout my life, in any public activity I may have engaged in, I am the same man as I am in private life. I have never come to an agreement with anyone to act unjustly, neither with anyone else nor with anyone of those who they slanderously say are my pupils. I have never been anyone's teacher. If anyone, young or old, desires to listen to me when I am talking and dealing with my own concerns, I have never begrudged this to anyone, but I do not converse when I receive a fee and not when I do not. I am equally ready to question the rich and the poor if anyone is willing to answer my questions and listen to what I say. And I cannot justly be held responsible for the good or bad conduct of these people, as I never promised to teach them anything and have not done so. If anyone says that he has learned anything from me, or that he heard anything privately that the others did not hear, be assured that he is not telling the truth.

Why then do some people enjoy spending considerable time in my company? You have heard why, gentlemen of the jury, I have told you the whole truth. They enjoy hearing those being questioned who think they are wise, but are not. And this is not unpleasant. To do this has, as I say, been enjoined c upon me by the god, by means of oracles and dreams, and in every other way that a divine manifestation has ever ordered a man to do anything. This is true, gentlemen, and can easily be established.

If I corrupt some young men and have corrupted others, then surely some of them who have grown older and realized that I gave them bad advice when they were young should now themselves come up here to accuse me and avenge themselves. If they were unwilling to do so themselves, then some of their kin- dred, their fathers or brothers or other relations should recall it now if their family had been harmed by me. I see many of these present here, first Crito, my contemporary and fellow demes- man, the father of Critoboulos here; next Lysanias of Sphet- tus, the father of Aeschines here; also Antiphon the Cephisian, the father of Epigenes; and others whose brothers spent their time in this way; Nicostratus, the son of Theozotides, brother of Theodotus, and Theodotus has died so he could not influ- ence him; Paralios here, son of Demodocus, whose brother was e Theages; there is Adeimantls, son of Ariston, brother of Plato here; Acantidorus, brother of Apollodorus here.

I could mention many others, some one of whom surely Meletus should have brought in as witness in his own speech. If he forgot to do so, then let him do it now; I will yield time if he has anything of the kind to say. You will find quite the 34 contrary, gentlemen. These men are all ready to come to the help of the corruptor, the man who has harmed their kindred, as Meletus and Anytus say. Now those who were corrupted might well have reason to help me, but the uncorrupted, their kindred who are older men, have no reason to help me except the right and proper one, that they know that Meletus is lying and that I am telling the truth.

Very well, gentlemen of the jury. This, and maybe other similar things, is what I have to say in my defence. Perhaps b one of you might be angry as he recalls that when he himself stood trial on a less dangerous charge, he begged and pleaded and implored the jury with many tears, that he brought his children and many of his friends and family into court to arouse as much pity as he could, but that I do none of these things, even though I may seem to be running the ultimate risk. Thinking of this, he might feel resentful toward me and, angry about this, cast his vote in anger. If there is such a one among you—I do c not deem there is, but if there is—I think it would be right to say in reply: My good sir, I too have a household and, in Homer's phrase, I am not born "from oak or rock" but from men, so that I have a family, indeed three sons, gentlemen of the jury, of whom one is an adolescent while two are children. Nevertheless, I will not beg you to acquit me by bringing them d here. Why do I do none of these things? Not through arrogance, gentlemen, nor through lack of respect for you. Whether I am brave in the face of death is another matter, but with regard to my reputation and yours and that of the whole city, it does not seem right to me to do these things, especially at my age and with my reputation. For it is generally believed, whether it be true or false, that in certain respects Socrates is superior to the majority of men. Now if those of you who are considered supe- e rior, be it in wisdom or courage or whatever other virtue makes them so, are seen behaving like that, it would be a disgrace. Yet I have often seen them do this sort of thing when standing trial, men who are thought to be somebody, doing amazing things as if they thought it a terrible thing to die, and as if they were to be immortal if you did not execute them. I think 35 these men bring shame upon the city so that a stranger, too, would assume that those who are outstanding in virtue among the Athenians, whom they themselves select from themselves to fill offices of state and receive other honours, are in no way better than women. You should not act like that, gentlemen of the jury, those of you who have any reputation at all, and if we

do, you should not allow it. You should make it very clear that you will more readily convict a man who performs these pitiful b  
dramatics in court and so makes the city a laughingstock, than a man who keeps quiet.

Quite apart from the question of reputation, gentlemen, I do not think it right to supplicate the jury and to be acquitted because of this but to teach and persuade them. It is not the purpose of a juryman's office to give justice as a favour to whoever seems good to him, but to judge according to law, and this c  
he has sworn to do. We should not accustom you to perjure yourselves, nor should you make a habit of it. This is irreverent conduct for either of us.

Do not deem it right for me, gentlemen of the jury, that I should act towards you in a way that I do not consider to be good or just or pious, especially, by Zeus, as I am being prosecuted by Meletus here for impiety; clearly, if I convinced you by my supplication to do violence to your oath of office, I d  
would be teaching you not to believe that there are gods, and my defence would convict me of hot believing in them. This is far from being the case, gentlemen, for I do believe in them as none of my accusers do. I leave it to you and the god to judge me in the way that will be best for me and for you.

[The jury now gives its verdict of guilty, and Meletus asks for the penalty of death.]

There are many other reasons for my not being angry with you for convicting me, gentlemen of the jury, and what happened was not unexpected. I am much more surprised at the number of votes cast on each side, for I did not think the decision would be by so few votes but by a great many. As it is, a switch of only thirty votes would have acquitted me. I think myself that I have been cleared on Meletus' charges, and e  
not only this, but it is clear to all that, if Anytus and Lycon had not joined him in accusing me, he would have been fined a 36  
thousand drachmas for not receiving a fifth of the votes.

He assesses the penalty at death. So be it. What counter-assessment should I propose to you, gentlemen of the jury? b

Clearly it should be a penalty I deserve, and what do I deserve to suffer or to pay because I have deliberately not led a quiet life but have neglected what occupies most people: wealth, household affairs, the position of general or public orator or the other offices, the political clubs and factions that exist in the city? I thought myself too honest to survive if I occupied myself with those things. I did not follow that path that would have made me of no use either to you or to myself, but I went to each of you privately and conferred upon him what I say is the greatest benefit, by trying to persuade him not to care for any of his belongings before caring that he himself should be as good and as wise as possible, not to care for the city's possessions more c than for the city itself, and to care for other things in the same way. What do I deserve for being such a man? Some good, gentlemen of the jury, if I must truly make an assessment according to my deserts, and something suitable. What is suitable for a poor benefactor who needs leisure to exhort you? Nothing is more suitable, gentlemen, than for such a man to be fed in the d Prytaneum, much more suitable for him than for anyone of you who has won a victory at Olympia<sup>5</sup> with a pair or a team of horses. The Olympian victor makes you think yourself happy; I make you be happy. Besides, he does not need food, but I do. So if I must make a just assessment of what I deserve, I assess it at this: free meals in the Prytaneum.

When I say this you may think, as when I spoke of appeals to pity and entreaties, that I speak arrogantly, but that is not the case, gentlemen of the jury; rather it is like this: I am convinced that I never willingly wrong anyone, but I am not convincing e you of this, for we have talked together but a short time. If 37 it were the law with us, as it is elsewhere, that a trial for life should not last one but many days, you would be convinced, but now it is not easy to dispel great slanders in a short time. Since I am convinced that I wrong no one, I am not likely to wrong myself, to say that I deserve some evil and to make some b such assessment against myself. What should I fear? That I should suffer the penalty Meletus has assessed against me, of



which I say I do not know whether it is good or bad? Am I then to choose in preference to this something that I know very well to be an evil and assess the penalty at that? Imprisonment? Why should I live in prison, always subjected to the ruling magistrates the Eleven? A fine, and imprisonment until I pay it? That would be the same thing for me, as I have no money. c Exile? for perhaps you might accept that assessment.

I should have to be inordinately fond of life, gentlemen of the jury, to be so unreasonable as to suppose that other men will easily tolerate my company and conversation when you, my fellow citizens, have been unable to endure them, but found them a burden and resented them so that you are now seeking to get rid of them. Far from it, gentlemen. It would be a fine life at my age to be driven out of one city after another, for d I know very well that wherever I go the young men will listen to my talk as they do here. If I drive them away, they will themselves persuade their elders to drive me out; if I do not drive them away, their fathers and relations will drive me out on their behalf. e

Perhaps someone might say: But Socrates, if you leave us will you not be able to live quietly, without talking? Now this is the most difficult point on which to convince some of you. If I say that it is impossible for me to keep quiet because that means disobeying the god, you will not believe me and will think I am being ironical. On the other hand, if I say that it is the greatest good for a man to discuss virtue every day and those other things about which you hear me conversing and testing myself and others, for the unexamined life is not worth living for man, you will believe me even less. 38

What I say is true, gentlemen, but it is not easy to convince you. At the same time, I am not accustomed to think that I deserve any penalty. If I had money. I would assess the penalty at the amount I could pay, for that would not hurt me, but I have none, unless you are willing to set the penalty at the amount I can pay, and perhaps I could pay you one mina of silver. So that is my assessment. b

Plato here, gentlemen of the jury, and Crito and Critobulus and Apollodorus bid me put the penalty at thirty minae, and they will stand surety for the money. Well then, that is my assessment, and they will be sufficient guarantee of payment.

[The jury now votes again and sentences Socrates to death.]

It is for the sake of a short time, gentlemen of the jury, that you will acquire the reputation and the guilt, in the eyes of those who want to denigrate the city, of having killed Socrates, a wise man, for they who want to revile you will say that I am wise even if I am not. If you had waited but a little while, this would have happened of its own accord. You see my age, that I am already advanced in years and close to death. I am saying c this not to all of you but to those who condemned me to death, and to these same jurors I say: Perhaps you think that I was convicted for lack of such words as might have convinced you, if I thought I should say or do all I could to avoid my sentence. Far from it. I was convicted because I lacked not words but boldness and shamelessness and the willingness to say to you d what you would most gladly have heard from me, lamentations and tears and my saying and doing many things that I say are unworthy of me but that you are accustomed to hear from others. I did not think then that the danger I ran should make me do anything mean, nor do I now regret the nature of my defence. I would much rather die after this kind of defence than live after making the other kind. Neither I nor any other man should, on trial or in war, contrive to avoid death at any cost. e Indeed it is often obvious in battle that one could escape death by throwing away one's weapons and by turning to supplicate one's pursuers, and there are many ways to avoid death in every kind of danger if one will venture to do or say anything to avoid 39 it. It is not difficult to avoid death, gentlemen of the jury, it is much more difficult to avoid wickedness, for it runs faster than death. Slow and elderly as I am, I have been caught by the slower pursuer, whereas my accusers, being clever and sharp, have been caught by the quicker, wickedness. I leave you now, condemned to death by you, but they are condemned by truth b

to wickedness and injustice. So I maintain my assessment, and they maintain theirs. This perhaps had to happen, and I think it is as it should be.

Now I want to prophesy to those who convicted me, for I am at the point when men prophesy most, when they are about to die. I say gentlemen, to those who voted to kill me, that vengeance will come upon you immediately after my death, a vengeance much harder to bear than that which you took in killing me. You did this in the belief that you would avoid giving an account of your life, but I maintain that quite the opposite will happen to you. There will be more people to test you, whom I now held back, but you did not notice it. They will be more difficult to deal with as they will be younger and you will resent them more. You are wrong if you believe that by killing people you will prevent anyone from reproaching you for not living in the right way. To escape such tests is neither possible nor good, but it is best and easiest not to discredit others but to prepare oneself to be as good as possible. With this prophecy to you who convicted me, I part from you.

I should be glad to discuss what has happened with those who voted for my acquittal during the time that the officers of the court are busy and I do not yet have to depart to my death. So, gentlemen, stay with me awhile, for nothing prevents us from talking to each other while it is allowed. To you, as being my friends, I want to show the meaning of what has occurred. A surprising thing has happened to me, judges—you I would rightly call judges. At all previous times my familiar prophetic power, my spiritual manifestation frequently opposed me, even in small matters, when I was about to do something wrong, but now that, as you can see for yourselves, I was faced with what one might think, and what is generally thought to be, the worst of evils, my divine sign has not opposed me, either when I left home at dawn, or when I came into court, or at any time that I was about to say something during my speech. Yet in other talks it often held me back in the middle of my speaking, but now it has opposed no word or deed of mine. What do I think

is the reason for this? I will tell you. What has happened to me may well be a good thing, and those of us who believe death to be an evil are certainly mistaken. I have convincing proof of this, for it is impossible that my familiar sign did not oppose me if I was not about to do what was right.

Let us reflect in this way, too, that there is good hope that death is a blessing, for it is one of two things: either the dead are nothing and have no perception of anything, or it is, as we are told, a change and a relocating for the soul from here to another place. If it is complete lack of perception, like a dreamless sleep, then death would be a great advantage. For I think that if one had to pick out that night during which a man slept soundly and did not dream, put beside it the other nights and days of his life, and then see how many days and nights had been better and more pleasant than that night, not only a private person but the great king would find them easy to count compared with the other days and nights. If death is like this I say it is an advantage, for all eternity would then seem to be no more than a single night. If, on the other hand, death is a change from here to another place, and what we are told is true and all who have died are there, what greater blessing could there be, gentlemen of the jury? If anyone arriving in Hades will have escaped from those who call themselves judges here, and will find those true judges who are said to sit in judgement there, Minos and Radamanthus and Aeacus and Triptolemus and the other demi-gods who have been upright in their own life, would that be a poor kind of change? Again, what would one of you give to keep company with Orpheus and Musaeus, Hesiod and Homer? I am willing to die many times if that is true. It would be a wonderful way for me to spend my time whenever I met Palamedes and Ajax, the son of Telamon, and any other of the men of old who died through an unjust conviction, to compare my experience with theirs. I think it would be pleasant. Most important, I could spend my time testing and examining people there, as I do here, as to who among them is wise, and who thinks he is, but is not.

What would one not give, gentlemen of the jury, for the opportunity to examine the man who led the great expedition b against Troy, or Odysseus, or Sisyphus, and innumerable other men and women one could mention. It would be an extraordinary happiness to talk with them, to keep company with them and examine them. In any case, they would certainly not put one to death for doing so. They are happier there than we c are here in other respects, and for the rest of time they are deathless, if indeed what we are told is true.

You too must be of good hope as regards death, gentlemen of the jury, and keep this one truth in mind, that a good man cannot be harmed either in life or in death, and that his affairs are not neglected by the gods. What has happened to me now has not happened of itself, but it is clear to me that it was better for me to die now and to escape from trouble. That is why my divine sign did not oppose me at any point. So I am certainly d not angry with those who convicted me, or with my accusers. Of course that was not their purpose when they accused and convicted me, but they thought they were hurting me, and for this they deserve blame. This much I ask from them: when my sons grow up, avenge yourselves by, causing them the same kind of grief that I caused you, if you think they care for money or anything else more than they care for virtue, or if they think they are somebody when they are nobody. Reproach them as e I reproach you, that they do not care for the right things and think they are worthy when they are not worthy of anything. If you do this, I shall have been justly treated by you, and my sons also.

Now the hour to part has come. I go to die, you go to live. Which of us goes to the better lot is known to no one, except the god.

# *Part 2: Life and Times of Socrates*

With the introduction to philosophy out of the way (as in, the barebones of the method and the quest for answers), we will now move on to the life and times of Socrates, the protagonist and real world person who you should be reading in the assigned reading for this module. This should serve as an ample starting point to understand the raise and fall of Philosophy's most famous proponent.

## **Birth and Parents**

Socrates was born around spring 469BCE in Athens. At that time, the Persians had just attempted (and failed) to invade Athens and Athens, on a global scale, was forming an alliance with the other city-states in the region (this alliance, called The Delian League, would grow into the Athenian Empire). This is on the Attic Peninsula and, politically and socially, it was divided into 139 districts, which were in turn, broken up among the 10 tribes which made up Athens. The members of these tribes were automatically Athenian. Socrates was a member of the Antiochis, which was located outside of the city walls, to the south-east. Keeping with the customs of the time,

5 days after Socrates' birth, his father, Sophroniscus, looked the infant Socrates over while walking around the hearth, to make sure that the kid was his, and then admitted him into the family. 5 days after this, Socrates was actually given a name and his father presented him to the local officials for the relevant paperwork (think of this as the equivalent of a birth-certificate and all that). In doing this, Sophroniscus took on the responsibility of ensuring that Socrates got a proper education and was made into a respectable member of society.

## Education

In Athens, the ability to read and write was commonplace since around 520BCE and it was unheard of for a young man not to have that skill. When Socrates turned 5, he started the equivalent of elementary school. His education consisted of learning to read and write, gymnastics (it was expected that Athenians be physically fit for military service, think of this like PE), music, and basic mathematics. It is because of this that Socrates, frankly, became a nerd. Socrates, as to be expected, loved philosophy. At that time, philosophy was, and still is, the mother of all other subjects, physics, mathematics, biology, and so on. The term 'philosophy' comes from two Greek roots, 'philein' meaning love and 'sophia' meaning wisdom. For Socrates and other Greeks (as well as any good philosophy/science today), philosophy was done with unaided reason and careful observation of the world. Over time, the careful observation aspect became what we call the sciences. But, to do philosophy, there are certain things which you don't get. You don't get anything in a religious scripture (so I don't want to see you reference a religious text in this class), you don't get what some authority figure says (just because they are an authority), you don't get myths, and you don't get things just because they were always done that way (no tradition/cultural beliefs). For philosophy, all you get is your observations and your own good reason.

At that time, philosophers were inventing Geometry, proposing a heliocentric model of the solar system (the Earth revolves around the sun), and also various aspects of the natural sciences which would develop into their own fields (EG biology).

Socrates, by all accounts, was a remarkably ugly person. Many accounts say that his childhood classmates would call him “frogface”. He had bulging eyes, broad, flat nose, and thick lips. He also walked bow-legged and sideways like a crab. If you ever see a bust of Socrates, understand that the carving is an angel compared to him in reality.

## ”Adult” Socrates

At the age of 17, Socrates graduated from the schools and was sworn in as a full-fledged citizen of Athens (think of it as like turning 18 in the US). Socrates’ father took him to the ceremony, but died shortly afterwards. His mother, Phaenarete, remarried and had another son, named Patrocles. In those days, Athens had one of the first constitutional democracies which enshrined freedom of speech and, because of this, public discussion, voting, and debate about all matters were quite commonplace. Athens had many festivals and gatherings throughout the year, attracting many of the great minds from around Greece. According to some accounts, at 19, Socrates was often found at these festivals discussing things with the philosophers of the day, allowable because of the freedom of speech. Freedom of speech is a definite perk which was not had by many societies in those days, but being an Athenian did come with some responsibilities. Upon becoming a citizen, Socrates was also put on the draft for military service.

## Military Career

After being called on for the draft and completing the required 2 years of training, Socrates served in the military for Athens.



During his first posting, it was a time of relative peace, so Socrates likely practiced a trade (stonecutting, like his father). But, as the years progressed, Athens was starting to get into a war with Sparta. Socrates served on the front lines for many of the battles in 432BCE, Socrates was called on to serve in the military for Athens on a few different occasions. After putting down a revolt, Socrates' troop entered into a very heavy conflict near Spartolus, where they suffered heavy casualties. In this fight, Socrates' bravery became legendary. He refused to retreat until he was the last person there and fought off enemy soldiers with another soldier, Alcibiades, on his back, saving his life. This deployment kept him away from home for 3 years. Socrates returned to duty again in 424BCE, where his bravery was again noted by the generals. His commander, Laches, noted Socrates' bravery when writing about the nature of courage, stating that he refused to retreat, even after the order was given, until he was the last person to leave. A year later, Socrates returned to duty and fought in another battle. After this, as far as we can tell, he did not serve in the military any longer. After this point, Athens and Sparta signed a treaty, which gave Athens a few years without the struggles of war. During this time, Socrates married Xanthippe, who, because of Socrates' military exploits, came with a large dowry.

## Family Man

Socrates cared little for material possessions. Often wearing the same clothes for many days in a row, including sleeping in them. While in the military, as well as throughout his life, Socrates rarely, if ever, wore footwear, even on ice or snow. Athenians of his day described him as "frugal". Although it was well within his means to afford various things, like shoes, he practiced "voluntary simplicity". Xanthippe gave birth to three sons, the third being born while Socrates was in prison awaiting execution, so they did the ceremony in the prison. According to

many accounts, Xanthippe had a volatile personality and was very unhappy with her hubby, and with good reason, he had a habit of spending all day and night in the agora (public market space) discussing questions and arguing with people when he could be out there making money (he was a stonecutter, like his father). According to one story, after chewing him out, Xanthippe climbed onto the roof of their house and dumped a bucket of urine on his head as Socrates went out to debate people.

Because of his willingness and constant engagement in questioning and debate, by his middle age, Socrates became a very recognizable (today we would say famous) person on the streets of Athens. In the Greek comedies of the day, an ugly caricature of Socrates was a re-occurring character. Socrates was made-fun of for his appearance and his love of philosophy in at least 3 prize-winning plays. Socrates took these in stride. At one point a foreigner asked “who’s this loon, Socrates?” in the middle of a play. At which point Socrates stood up and said “ME!”

## Part 2.1: Trials in Athens

### How Trials Worked in Athens

In Socrates’ time, the procedure for a trial and court cases like this was pretty well established. In that time, also, there was no such thing as what we would call today a ‘public prosecutor’, this is similar to how in London, prior to their police force, they had a hew and cry system. If a citizen suspected another of a crime, then they would report it to the officials to have the case looked over. Any actual citizen of Athens could initiate the procedure. Once such a crime was reported, the trial consisted of three parts. First, there was the Initiation of Criminal Proceedings. Next, they had The Preliminary Hearing (Anakrisis). And third, they had The Trial itself. As evidence for how common and stream-lined this process was in Athens, historically,

just prior to his trial, Socrates was engaged in a conversation with a prosecutor for a trial just ending. This conversation was on the nature of goodness and the question related to it is the discussion for this module. However, most cases were settled prior to reaching the actual trial because, as we will see, making the arrangements is quite the undertaking.

## **Initiation of Criminal Proceedings**

As I just mentioned, any Athenian could raise charges against another. In the case of Socrates, the proceedings began when Meletus, a poet, arranged a meeting, for a specific date, with the legal magistrate, or in some cases King Archon, in a colonnaded building called the Royal Stoa to answer charges of impiety. Meletus, then delivered an oral summons to Socrates in the presence of witnesses (or callers). Once the magistrate determined – after listening to Socrates and Meletus (and perhaps the other two accusers, Anytus and Lycon) – that the lawsuit was permissible under Athenian law, a date was set for the “preliminary hearing” (anakrisis) and terms for the hearing were posted as a public notice at the Royal Stoa.

## **The Preliminary Hearing (Anakrisis)**

The preliminary hearing before the magistrate at the Royal Stoa began with the reading of the written charge by Socrates’ accuser, Meletus. Socrates then formally answered the charge. Both the written charge and denial were then attested to by each, under oath, as being true. The next phase of the preliminary hearing was an interrogation. First, the magistrate questioned both Meletus and Socrates. This is to see whether the issue could be settled out of court and to see whether there was any merit to the charges. Second, both the accuser and defendant were allowed to question each other. This was another chance to change the course of events. The third and final phase, supposing that the magistrate found the case worthy of

consideration, the magistrate would draw up formal charges against the accused and set a date for the public trial. For Socrates, these charges (relating to impiety and corruption of youth), the actual paperwork, were preserved as a public document (antomosia), and they survived until at least the second century C.E., but were subsequently lost.

## The Trial

The trial of Socrates took place over a nine-to-ten hour period in the People's Court, located in the agora, the civic center of Athens. The jury consisted of 500 male citizens over the age of thirty, chosen by lot from among volunteers. People today might think that this is a ridiculously high number for a jury. But, sometimes, the juries could be as large as 1501 men. This was a protection against bribery. For example, in order to ensure that a case went a certain way, the person would need to bribe at least 251 people, and it's not likely that many could afford to do this. All jurors were required to swear by the gods of Zeus, Apollo, and Demeter the Heliastic Oath: "I will cast my vote in consonance with the laws and decrees passed by the Assembly and by the Council, but, if there is no law, in consonance with my sense of what is most just, without favor or enmity. I will vote only on the matters raised in the charge, and I will listen impartially to the accusers and defenders alike."

Most of the jurors were probably farmers, as that was the principal occupation of the day. For their jury service they received payment of three obols. An obol was a currency of the day and it was around a 40TH of an ounce of silver. In general, three obols fed you and paid for a leisurely night. The jurors sat on wooden benches separated from spectators by some sort of barrier or railing. Given Socrates's fame and the notoriousness of the charge against him, the crowd of spectators was most likely large – including, of course, the most famous pupil of Socrates, Plato.

The trial began in the morning with the reading of the for-

mal charges against Socrates by a herald. Few, if any, formal rules of evidence existed. The prosecution presented its case first. Meletus, Anytus, and Lycon had three hours, measured by a waterclock, to make their argument for a finding of guilt. Each accuser spoke from an elevated stage. No record of the prosecution's argument against Socrates survives. Following the prosecution's case, Socrates had three hours to answer the charges. Although many written versions of the defense – or apology – of Socrates at one time circulated, only two have survived: one by Plato and another by Xenophon. After the arguments, the herald of the court called on the jurors to consider their decision. In Athens, jurors did not retire to a juryroom to deliberate – they made their decisions without discussion among themselves, based in large part on their own interpretations of the law. The 500 jurors voted on his guilt or innocence by dropping bronze ballot disks into marked urns. Only a majority vote was necessary for conviction. Four jurors were assigned the task of counting votes. In the case of Socrates, the jury found him guilty on a relatively close vote of 280 to 220. (Interestingly, if less than 100 jurors voted for guilt, the accusers had to pay a fine to cover trial costs, which is similar to something which is found in court cases today.)

## The Final Phase

If a defendant is convicted, the trial enters a second phase to set punishment. The prosecution and the defendant each propose a punishment and the jury chooses between the two punishment options presented to it. The range of possible punishments included death, imprisonment, loss of civil rights (i.e., the right to vote, the right to serve as a juror, the right to speak in the Assembly), exile, and fines. In the trial of Socrates, the principal accusers proposed the punishment of death. Socrates, if Plato's account is to be believed, proposed first the punishment – or, rather, the nonpunishment – of free meals in the center of the city, then later the extremely modest fine of one mina of silver.

Apparently finding Socrates' proposed punishment insultingly light, the jury voted for the prosecution's proposal of death by a larger margin than for conviction, 360 to 140. The execution of Socrates was accomplished through the drinking of a cup of poison hemlock.

## Part 2.2: Socrates' Defense/Arguments

### The Charges

There were three formal charges which Socrates was going to need to reply to in his 3 hour block of time. The first, the real zinger, was corrupting the youth of Athens. This was essentially causing the youth to have rebellious mindsets and be against the state. There was some evidence to this claim as several attempted rebellions and revolutions had taken place with many of the young people which Socrates spoke with often playing central roles, though Socrates himself was more of a bystander than anything in these cases. The second charge, inventing new gods, was a lesser case. This is more in line with the initial charge of impiety. According to the laws of Athens, blasphemy of this sort was a crime punishable by death. The third charge, which could have either built up or dismantled the second, was not believing in the gods of Athens. This could have gone two ways, either the accusers could have said that in inventing new gods, Socrates abandoned belief in the Athenian pantheon (which would have built up the case regarding blasphemy) or they could have said that Socrates was an atheist, which directly contradicts the second charge. As it turns out, the accusers, not trained in philosophy, did not see the contradiction and went with the latter option (being an atheist).

## Socrates' Method

Philosophy, in this period, was still very much in its infancy, it wasn't as formalized as it is today (throughout this class, you will be learning the philosophic method, which is the great and powerful grandmother to the scientific method which you may be familiar with). For the pre-socratics (which includes Socrates), you can think of philosophy as a game where they are still trying to figure out the rules. If you have ever seen little kids try and play soccer or football for the first time, then you will have the right kind of image in mind. Socrates, for his part, in the trial, was unfamiliar and inexperienced in how to present publicly, especially given the large number of people in attendance, and was not really the best at being persuasive (rhetoric was an area which he really did not have much respect for). So, Socrates begins his rebuttal by explaining that he will treat this like he would a debate in the Agora. In those cases, Socrates would merely ask questions to people. Rarely, if ever, does Socrates actually state his stance. Rather, Socrates merely restates the stance which the person had stated in reply to his question and then adds further questions to this. If you ever get the chance to see me in a lecture format, I do this methodology for the first few days of lecture, off and on. As a result, Socrates did not address the jury directly, rather just questioned Meletus. Though this method is quite good for teaching and one-on-one debate, it's not the best for a presentation of this sort.

## Socrates' Questions

Since I understand that the translations for this are a bit wordy and the word choice can be a bit unnatural (I have a lot of experience translating Latin and if a person is not careful or doesn't think about the context of the wording, the translation will come out as quite formal), I have taken the liberty of rewording

some of the content of The Apology to make the meaning come through a little clearer. It's still important that you read the original, as it's a wonderful work, however, this is here to help. For this, you can read it as a script, with 'S' being Socrates and 'M' being Meletus.

## Corrupting the Youth

This charge, as I have mentioned, is the real big charge, the others, though still punishable by death, don't have as much going for them, as we will see. To prove his point, Socrates asked Meletus, around, 11 questions, which were leading him down a very particular rabbit-hole.

S: Do you think that it is the greatest importance to make the youth as great as possible?

M: I do.

S: Who improves them?

M: The Laws

S: You didn't answer me, Who has knowledge of these laws?

M: These Jurymen

S: What do you mean? Do all these people improve the youth?

M: All of them.

So, at this point, Meletus starts off by essentially claiming that the laws of Athens are the sort of things which can improve the youth. But this is not correct, laws are passive actors, they aren't the sort of thing which can take an active part in the raising of the youth. That requires a person who knows the laws and has the ability to accurately teach. So, to his credit, Meletus noticed this and backtracked, claiming that the jurymen have the knowledge of the laws and the ability to teach them (think about how it would have sounded if he claimed that the jurors didn't know the laws).



S: By Hera! That's a lot of people to improve our youth, what great shape we must be in! What about the audience? So they improve them?

M: They do.

S: What about the council?

M: They improve them too.

S: Everyone in Athens makes the youth better but me, is that what you are saying?

M: Yep, everyone makes the youth better but you.

At this point, Meletus doesn't seem to see where this is going. Socrates has, through his questions, basically forced Meletus to claim that all the people of Athens improve the youth, with Socrates being the sole bad influence. From a rhetorical side, this was an interesting ploy. If Meletus had claimed otherwise, then Socrates could have easily countered by asking why those people weren't on trial with him, why charges hadn't be posed against them. But, since Meletus said that Socrates was the only corrupter, Socrates needs to go a different route, which leads us to the final 'build-up' question:

S: Man, you put me in a rough spot. But does this also apply to horses? That all people improve them and only one person corrupts them? Obviously, it is the opposite. The horse breeder (trainer) improves them and all others corrupt them. It would be a wonderful world if our youth only had one corrupting influence in their lives!

The horse analogy has not aged well, as for contemporary English, calling a person a horse is a bit insulting. However, think about it this way, a good horse was like the nice car of its day. The vast majority of people aren't trained in how to repair and maintain a car, or in this case, care for and train a horse. Rather, there are very few people who really know how to maintain those things, others who use them, rather uninten-

tionally, corrupt them (wear them out, make them forget their training, etc.). In the case of the youth, the vast majority of people aren't trained in the best child-raising techniques, they interact with children in unintentional ways which lead to negative behaviors, and so forth. It would be a great world if the children only had one corrupting influence, as that would get quickly drowned out by all of the other influences around them.

This leads us to the final set of questions concerning this particular accusation.

S: You have made it clear that you care very little for the youth and haven't put good thought into the reason you're putting me on trial. Now answer me this: Is it better for a person to live among good fellows or wicked? This isn't a hard question. Do wicked people harm their neighbors and good people benefit them?

M: Certainly

S: Now, is there a person alive who would rather be harmed than benefited?

M: Certainly not

S: Also, do you accuse me of corrupting the youth deliberately or unintentionally?

M: Deliberately

S: What then, Meletus? Are you so wise that you realize that the wicked do wicked and the good good but I am so stupid that I have not realized this? Have I failed to recognize that if I make anyone of my associates bad, they will harm me? But still you say that I do this great evil voluntarily? I can't believe you, Meletus, nor do I think anyone else in the world does! But, leaving that aside, either I do not corrupt them, or if I corrupt them, I do it involuntarily. This shows that you are lying either way we go. But if I corrupt them involuntarily, the law is not to haul people into court, but to take them and instruct and admonish them in private. For it is clear that if I am told about it, I shall stop doing that which I do involuntarily. But you avoided associating with me and instructing me, and were unwilling to do so, but you haul me in here, where it is the law to haul in those who need punishment, not instruction.

This is where we get Socrates' real argument against the charge. It follows a very interesting logical form, called Constructive Dilemma. It is one which Socrates, historically, had just used earlier in the day outside of the court concerning the

nature of goodness. In an abstract form, the reasoning goes like this: One of two (or more) options is correct (A or B). If A is correct, then C and if B is correct, then D. Therefore, either C or D is correct. Here is the argument for the case along with another example used by Socrates in another dialogue:

#### Corrupting the Youth

Either I don't corrupt the youth or I corrupt them unintentionally.

If I don't corrupt the youth, then the charge against me is false (I'm innocent here).

If I corrupt the youth unintentionally, then you should have instructed me in private and not taken me to court (I'm innocent here).

So, either way, I am innocent of this charge.

#### Piety

Either something is moral because it's commanded by the gods or it's commanded by the gods because it's moral.

If it's moral because it's commanded by the gods, then the morality of actions is arbitrary (random, not fixed).

If it's commanded by the gods because it's moral, then the morality of actions is not determined by the gods (we don't need to reference them to figure out the morality of actions).

So, either morality is arbitrary or it's not determined by the gods.

The argument regarding piety can also work if you replace 'the gods' with 'God'. If you are interested in that particular argument, look into the Euthyphro Dilemma.

## MODULE II

*Is Death Bad For The  
Person Who Died?*

# *Death By Thomas*

## *Nagel*

19

If death is the unequivocal and permanent end of our existence, the question arises whether it is a bad thing to die.

There is conspicuous disagreement about the matter: some people think death is dreadful; others have no objection to death per se, though they hope their own will be neither premature nor painful. Those in the former category tend to think those in the latter are blind to the obvious, while the latter suppose the former to be prey to some sort of confusion. On the one hand it can be said that life is all we have and the loss of it is the greatest loss we can sustain. On the other hand it may be objected that death deprives this supposed loss of its subject, and that if we realize that death is not an unimaginable condition of the persisting person, but a mere blank, we will see that it can have no value whatever, positive or negative.

Since I want to leave aside the question whether we are, or might be, immortal in some form, I shall simply use the word 'death' and its cognates in this discussion to mean permanent death, unsupplemented by any form of conscious survival. I

---

<sup>19</sup>Thomas Nagel, "Death," *Noûs* vol. 4, no. 1. *JSTOR*, 1970, [www.jstor.org/stable/2214297](http://www.jstor.org/stable/2214297): pp. 73–80.

want to ask whether death is in itself an evil; and how great an evil, and of what kind, it might be. The question should be of interest even to those who believe in some form of immortality, for one's attitude towards immortality must depend in part on one's attitude toward death.

If death is an evil at all, it cannot be because of its positive features, but only because of what it deprives us of. I shall try to deal with the difficulties surrounding the natural view that death is an evil because it brings to an end all the goods that life contains. We need not give an account of these goods here, except to observe that some of them, like perception, desire, activity, and thought, are so general as to be constitutive of human life. They are widely regarded as formidable benefits in themselves, despite the fact that they are conditions of misery as well as of happiness, and that a sufficient quantity of more particular evils can perhaps outweigh them. That is what is meant, I think by the allegation that it is good simply to be alive, even if one is undergoing terrible experiences. The situation is roughly this: There are elements which, if added to one's experience, make life better; there are other elements which if added to one's experience, make life worse. But what remains when these are set aside is not merely neutral: it is emphatically positive. Therefore life is worth living even when the bad elements of experience are plentiful, and the good ones too meager to outweigh the bad ones on their own. The additional positive weight is supplied by experience itself, rather than by any of its consequences. I shall not discuss the value that one person's life or death may have for others, or its objective value, but only the value that it has for the person who is its subject. That seems to me the primary case, and the case which presents the greatest difficulties. Let me add only two observations. First, the value of life and its contents does not attach to mere organic survival; almost everyone would be indifferent (other things equal) between immediate death and immediate coma followed by death twenty years later without reawakening. And second, like most goods, this can be mul-

tiplied by time: more is better than less. The added quantities need not be temporally continuous (though continuity has its social advantages). People are attracted to the possibility of long-term suspended animation or freezing, followed by the resumption of conscious life, because they can regard it from within simply as a continuation of their present life. If these techniques are ever perfected, what from outside appeared as a dormant interval of three hundred years could be experienced by the subject as nothing more than a sharp discontinuity in the character of his experiences. I do not deny, or course, that this has its own disadvantages. Family and friends may have died in the meantime; the language may have changed; the comforts of social, geographical, and cultural familiarity would be lacking. Nevertheless those inconveniences would not obliterate the basic advantage of continued, though discontinuous, existence.

If we turn from what is good about life to what is bad about death, the case is completely different. Essentially, though there may be problems about their specification, what we find desirable in life are certain states, conditions, or types of activity. It is being alive, doing certain things, having certain experiences, that we consider good. But if death is an evil, it is the loss of life, rather than the state of being dead, or nonexistent, or unconscious, that is objectionable.<sup>20</sup> This asymmetry is important. If it is good to be alive, that advantage can be attributed to a person at each point of his life. It is a good of which Bach had more than Schubert, simply because he lived longer. Death, however, is not an evil of which Shakespeare has so far received a larger portion than Proust. If death is a disadvantage, it is not easy to say when a man suffers it.

There are two other indications that we do not object to death merely because it involves long periods on nonexistence. First, as has been mentioned, most of us would not regard the temporary suspension of life, even for substantial intervals, as

---

<sup>20</sup> It is often said that those who object to death have made the mistake of trying to imagine what it is like to be dead.



in itself a misfortune. If it ever happens that people can be frozen without reduction of the conscious lifespan, it will be inappropriate to pity those who are temporarily out of circulation. Second, none of us existed before we were born (or conceived), but few regard that as a misfortune. I shall have more to say about this later.

The point that death is not regarded as an unfortunate state enables us to refute a curious but very common suggestion about the origin of the fear of death. It is alleged that the failure to realize that this task is logically impossible (for the banal reason that there is nothing to imagine) leads to the conviction that death is mysterious and therefore a terrifying prospective state. But this diagnosis is evidently false, for it is just as impossible to imagine being totally unconscious as to imagine being dead (though it is easy enough to imagine oneself, from the outside, in either of those conditions). Yet people who are averse to death are not usually averse to unconsciousness (so long as it does not entail a substantial cut in the total duration of waking life).

If we are to make sense of the view that to die is bad, it must be on the ground that life is a good and death is the corresponding deprivation or loss, bad not because of any positive features but because of the desirability of what it removes. We must now turn to the serious difficulties which this hypothesis raises, difficulties about loss and privation in general, and about death in particular.

Essentially, there are three types of problem. First, doubt may be raised whether anything can be bad for a man without being positively unpleasant to him: specifically, it may be doubted that there are any evils which consist merely in the deprivation or absence of possible goods, and which do not depend on someone's minding that deprivation. Second, there are special difficulties, in the case of death, about how the supposed misfortune is to be assigned to a subject at all. There is doubt both to who its subject is, and as to when he undergoes it. So long as a person exists, he has not yet died, and once he has

died, he no longer exists; so there seems to be no time when death, if it is a misfortune, can be ascribed to its unfortunate subject. The third type or difficulty concerns the asymmetry, mentioned above, between our attitudes to posthumous and prenatal nonexistence. How can the former be bad if the latter is not?

It should be recognized that if these are valid objections to counting death as an evil, they will apply to many other supposed evils as well. The first type of objection is expressed in general form by the common remark that what you don't know can't hurt you. It means that even if a man is betrayed by his friends, ridiculed behind his back, and despised by people who treat him politely to his face, none of it can be counted as a misfortune for him so long as he does not suffer as a result. It means that a man is not injured if his wishes are ignored by the executor of his will, or if, after his death, the belief becomes current that all the literary works on which his fame rest were really written by his brother, who died in Mexico at the age of 28. It seems to me worth asking what assumptions about good and evil lead to these drastic restrictions.

All the questions have something to do with time. There certainly are goods and evils of a simple kind (including some pleasures and pains) which a person possesses at a given time simply in virtue of his condition at that time. But this is not true of all the things we regard as good or bad for a man. Often we need to know his history to tell whether something is a misfortune or not; this applies to ills like deterioration, deprivation, and damage. Sometimes his experiential state is relatively unimportant – as in the case of a man who wastes his life in the cheerful pursuit of a method of communicating with asparagus plants. Someone who holds that all goods and evils must be temporally assignable states of the person may of course try to bring difficult cases into line by pointing to the pleasure or pain that more complicated goods and evils cause. Loss, betrayal, deception, and ridicule are on this view bad because people suffer when they learn of them. But it

should be asked how our ideas of human value would have to be constituted to accommodate these cases directly instead. One advantage of such an account might be that it would enable us to explain why the discovery of these misfortunes causes suffering – in a way that makes it reasonable. For the natural view is that the discovery of betrayal makes us unhappy because it is bad to be betrayed – not that betrayal is bad because its discovery makes us unhappy.

It therefore seems to me worth exploring the position that most good and ill fortune has as its subject a person identified by his history and his possibilities, rather than merely by his categorical state of the moment – and that while this subject can be exactly located in a sequence of places and times, the same is not necessarily true of the goods and ills that befall him.<sup>21</sup>

These ideas can be illustrated by an example of deprivation whose severity approaches that of death. Suppose an intelligent person receives a brain injury that reduces him to the mental condition of a contented infant, and that such desires as remain to him can be satisfied by a custodian, so that he is free from care. Such a development would be widely regarded as a severe misfortune, not only for his friends and relations, or for society, but also and primarily, for the person himself. This does not mean that a contented infant is unfortunate. The intelligent adult who has been reduced to this condition is the subject of the misfortune. He is the one we pity, though of course he does not mind his condition. It is in fact the same condition he was in at the age of three months, except that he is bigger. If we did not pity him then, why pity him now; in any case, who is there to pity? The intelligent adult has disappeared, and for a creature like the one before us, happiness consists in a full stomach and a dry diaper.

If these objections are invalid, it must be because they rest

---

<sup>21</sup>t is certainly not true in general of the things that can be said of him. For example, Abraham Lincoln was taller than Louis XIV. But when?

on a mistaken assumption about the temporal relation between the subject of a misfortune and the circumstances which constitute it. If, instead of concentrating exclusively on the oversized baby before us, we consider the person he was, and the person he could be now, then his reduction to this state and the cancellation of his natural adult development constitute a perfectly intelligible catastrophe.

This case should convince us that it is arbitrary to restrict the goods and evils that can befall a man to nonrelational properties ascribable to him at particular times. As it stands, that restriction excludes not only such cases of gross degeneration, but also a good deal of what is important about success and failure, and other features of a life that have the character of processes. I believe we can go further, however. There are goods and evils which are irreducibly relational; they are features of the relations between a person, with spatial and temporal boundaries of the usual sort, and circumstances which may not coincide with him either in space or in time. A man's life includes much that does not take place within the boundaries of his life. These boundaries are commonly crossed by the misfortunes of being deceived, or despised, or betrayed. (If this is correct, there is a simple account of what is wrong with breaking a deathbed promise. It is an injury to the dead man. For certain purposes it is possible to regard time as just another type of distance.). The case of mental degeneration shows us an evil that depends on a contrast between the reality and the possible alternatives. A man is the subject of good and evil as much because he has hopes which may or may not be fulfilled, or possibilities which may or may not be realized, as because of his capacity to suffer and enjoy. If death is an evil, it must be accounted for in these terms, and the impossibility of locating it within life should not trouble us.

When a man dies we are left with his corpse, and while a corpse can suffer the kind of mishap that may occur to an article of furniture, it is not a suitable object for pity. The man, however, is. He has lost his life, and if he had not died, he

would have continued to live it, and to possess whatever good there is in living. If we apply to death the account suggested for the case of dementia, we shall say that although the spatial and temporal locations of the individual who suffered the loss are clear enough, the misfortune itself cannot be so easily located. One must be content just to state that his life is over and there will never be any more of it. That fact, rather than his past or present condition, constitutes his misfortune, if it is one. Nevertheless if there is a loss, someone must suffer it, and he must have existence and specific spatial and temporal location even if the loss itself does not. The fact that Beethoven had no children may have been a cause of regret to him, or a sad thing for the world, but it cannot be described as a misfortune for the children that he never had. All of us, I believe, are fortunate to have been born. But unless good and ill can be assigned to an embryo, or even to an unconnected pair of gametes, it cannot be said that not to be born is a misfortune. (That is a factor to be considered in deciding whether abortion and contraception are akin to murder.)

This approach also provides a solution to the problem of temporal asymmetry, pointed out by Lucretius. He observed that no one finds it disturbing to contemplate the eternity preceding his own birth, and he took this to show that it must be irrational to fear death, since death is simply the mirror image of the prior abyss. That is not true, however, and the difference between the two explains why it is reasonable to regard them differently. It is true that both the time before a man's birth and the time after his death is time of which his death deprives him. It is time in which, had he not died then, he would be alive. Therefore any death entails the loss of some life that its victim would have led had he not died at that or any earlier point. We know perfectly well what it would be for him to have had it instead of losing it, and there is no difficulty in identifying the loser.

But we cannot say that the time prior to a man's birth is time in which he would have lived had he been born not

then but earlier. For aside from the brief margin permitted by premature labor, he could not have been born earlier: anyone born substantially earlier than he would have been someone else. Therefore the time prior to his birth prevents him from living. His birth, when it occurs, does not entail the loss to him of any life whatever.

The direction of time is crucial in assigning possibilities to people or other individuals. Distinct possible lives of a single person can diverge from a common beginning, but they cannot converge to a common conclusion from diverse beginnings. (The latter would represent not a set of different possible lives of one individual, but a set of distinct possible individuals, whose lives have identical conclusions.) Given an identifiable individual, countless possibilities for his continued existence are imaginable, and we can clearly conceive of what it would be for him to go on existing indefinitely. However inevitable it is that this will not come about, its possibility is still that of the continuation of a good for him, if life is the good we take it to be.<sup>22</sup>

---

<sup>22</sup>I confess to being troubled by the above argument, on the ground that it is too sophisticated to explain the simple differences between our attitudes to prenatal and posthumous nonexistence. For this reason I suspect that something essential is omitted from the account of the badness of death by an analysis which treats it as a deprivation of possibilities. My suspicion is supported by the following suggestion of Robert Nozick. We could imagine discovering that people developed from individual spores that had existed indefinitely far in advance of their birth. In this fantasy, birth never occurs naturally more than a hundred years before the permanent end of the spore's existence. But then we discover a way to trigger the premature hatching of these spores, and people are born who have thousands of years of active life before them. Given such a situation, it would be possible to imagine oneself having come into existence thousands of years previously. If we put aside the question whether this would really be the same person, even given the identity of the spore, then the consequence appears to be that a person's birth at a given time could deprive him of many earlier years of possible life. Now while it would be cause for regret that one had been deprived of all those possible years of life by being born too late, the feeling would differ from that which many people have about death. I conclude that something about the future prospect of permanent nothingness is not captured by the analysis in terms of denied possibilities.

We are left, therefore, with the question whether the non-realization of this possibility is in every case a misfortune, or whether it depends on what can naturally be hoped for. This seems to me the most serious difficulty with the view that death is always an evil. Even if we can dispose of the objections against admitting misfortune that is not experienced, or cannot be assigned to a definite time in the person's life, we still have to set some limits on how possible a possibility must be for its nonrealization to be a misfortune (or good fortune, should the possibility be a bad one). The death of Keats at 24 is generally regarded as tragic; that of Tolstoy at 82 is not. Although they will both be dead forever, Keats' death deprived him of many years of life which were allowed to Tolstoy; so in a clear sense Keats' loss was greater (though not in the sense standardly employed in mathematical comparison between infinite quantities). However, this does not prove that Tolstoy's loss was insignificant. Perhaps we record an objection only to evils which are gratuitously added to the inevitable; the fact that it is worse to die at 24 than at 82 does not imply that it is not a terrible thing to die at 82, or even at 806. the question is whether we can regard as a misfortune any limitations, like mortality, that is normal to the species. Blindness or near-blindness is not a misfortune for a mole, nor would it be for a man, if that were the natural condition of the human race.

The trouble is that life familiarizes us with the goods of which death deprives us. We are already able to appreciate them, as a mole is not able to appreciate vision. If we put aside doubts about their status as goods and grant that their quantity is in part a function of their duration, the question remains whether death, no matter when it occurs, can be said to deprive its victim of what is in the relevant sense a possible

---

If so, then Lucretius' argument still awaits an answer. I suspect that it requires a general treatment of the difference between past and future in our attitudes toward our own lives. Our attitudes toward past and future pain are very different, for example. Derek Parfit's unpublished writings on this topic have revealed its difficulty to me.

continuation of life.

The situation is an ambiguous one. Observed from without, human beings obviously have a natural lifespan and cannot live much longer than a hundred years. A man's sense of his own experience, on the other hand, does not embody this idea of a natural limit. His existence defines for him an essentially open-ended possible future, containing the usual mixture of goods and evils that he has found so tolerable in the past. Having been gratuitously introduced to the world by a collection of natural, historical, and social accidents, he finds himself the subject of a life, with an indeterminate and not essentially limited future. Viewed in this way, death, no matter how inevitable, is an abrupt cancellation of indefinitely extensive possible goods. Normality seems to have nothing to do with it, for the fact that we will all inevitably die in a few score years cannot by itself imply that it would be good to live longer. Suppose that we were all inevitably going to die in agony – physical agony lasting six months. Would inevitability make that prospect any less unpleasant? And why should it be different for a deprivation? If the normal lifespan were a thousand years, death at 80 would be a tragedy. As things are, it may just be a more widespread tragedy. If there is no limit to the amount of life that it would be good to have, then it may be that a bad end is in store for us all.



## *Part 3: How Does*

## *Nagel Define Death?*

In the reading, Nagel argues from the perspective that death is just the end of conscious experience, after that point, there are no feelings, no emotions, no thought for the person who died. One could think that there is some kind of after-life; if you think that, then you could think of this as a ‘what-if’ sort of mindset. This could be either a best-case or worst-case scenario. The question which Nagel is looking into is whether death, as we have so defined it, is bad for the person who died. I am not asking whether it is bad for the family or loved ones, as they will certainly be sad and that is bad, but what about the person who died? Is it bad for them?

For some practice in abstract thinking, think about how Nagel is defining death. Could there be cases where a human is dead, by Nagel’s definition, but is technically still alive? (Give this a moment’s thought before continuing.) One such example could be a case of a human, who was previously perfectly normal with all of the relevant faculties, entering into an irreversible coma. They would be, to use a common phrase, brain-dead. Their heart would still be beating and maybe they would still be breathing (we can easily assume that they could need a breathing machine), so they would still be alive, but without

the ability to have or process experiences, they would be dead. There could be other examples like this, but you need to be careful because it needs to be the permanent end of conscious experience (so falling into a deep sleep doesn't count).

So, if death is bad for the person who died, it can't be because of what it has (because death, as defined here, is the absence of everything). If it's bad, it must be because of what it lacks. Nagel lists four (4) different generic 'things' which living has but death lacks.

First, we have perception. Perception is the ability to take in and process stimuli. This is not necessarily the stimuli we get from the external world, like sight, sound, smell, and touch, but also includes the stimuli we get internally from us thinking. Death, since it doesn't have consciousness, can't have any perception (either internal or external).

Second, we have thought. Thinking is essential to the ability to perceive the world, either internal or external (though, it could be argued, easily, that perception is essential to thought). In death, we can't access memories, think about our lives, engage in puzzles, or anything. There is nothing.

The third is desire. We all have things which we want. The feeling of wanting something is desire. Though this requires thought, there are additional features to desire. Without desire, we can't love, hope, plan, go on adventures, or have fun (just to name a few things). Without desire, also, we can't experience heart-break, disappointment, frustration, or boredom. Death can't have these.

The final example of what death lacks is activity. Activity seems to require all of the previous (at least activity which you want to do requires desire). In death, we can't enact our plans, fulfill our desires, or do anything.

Next, we will move on to why Nagel thinks death is bad for the person who died. This is a common confusion for students. Nagel is arguing that death is bad for the person who died.

## Part 3.1: Is Death Bad for The Person Who Died?’

As I have mentioned previously, if death is bad, then it must be because of what it lacks and according to Nagel, death lacks four (4) generic things, perception, thought, desire, and activity. So, the first question Nagel needs to grapple with is whether some state of affairs can be bad (for you) because it causes you to miss out on something (in the case of death, this would be those for things).

Suppose that you live in an apartment above the only pizzeria in town. At the end of the day, you get the left over pizza. But, one day, that pizzeria shuts down. When it shuts down, not only will you miss out on the good of having a place to live but you will also miss out on the good of the greasy nummy cheesy pizza.

If you think that the pizza shop closing down was bad for you, then you are thinking like Nagel in this case (I could easily substitute different examples for this). The key thing is that there was a good which you were getting, and now you aren't, so you are worse off because of the lack. If you think about it, the four things which Nagel lists are the four things necessary for anything to be good for us. So if you die, you lose out on the good of the ability to experience goods. The lack of any potential good would seem to make death bad for the person who died.

### The Problem and Nagel's Take

The problem here is that the very same things which are required to experience good things in the world (perception, thought, desire, and activity) are the things which are required to experience bad. For example, without perception, you can't experience pain; without thought, you can't experience sadness;

without desire, you can't experience disappointment; without activity, you can't experience frustration. There is no aspect of experience which can't be used to have a bad one. It would seem from this that there can be some sets of experiences, with no end in sight, which can make life not worth living any longer. For example, a person with a terminal illness slowly and painfully killing them. How much painful or bad experiences does a person have to go through before life is not worth living any longer?

Nagel has a reply to this because he wants to say that death is always bad for the person who died. He says that experience itself is always good. You can think of experience itself as a variable value which is always just enough to make the value of some experience better than not having any at all. Yes, some experiences are better than others, but any experience is better than not having any at all, according to Nagel. Some of you might have heard the phrase "well, at least you experienced it", that's the sort of intuition Nagel is coming from. So, in response to the terminal illness, Nagel would say that it's a bad experience, but it's better than having none at all.

## **Part 3.2 The Three Objections to Nagel's Stance**

Nagel now turns to discussing the objections others might have to his stance. This is fairly common practice in Philosophy. You look at your position and think about how another person, on the opposing side, would combat or object to your thinking. Doing this makes you prepared for the upcoming debate. Nagel seeks to handle three possible objections to his stance. This is the bulk of the paper and the main topic for the homework this module.

## Can anything be bad for a person if it is not unpleasant to them?

Another way of phrasing this would be to ask whether there's anything which is bad merely because we miss out on it or are ignorant of it. This line of thinking is akin to the phrase "what you don't know can't hurt you." Nagel says that there is a problem with this objection, namely, that if it applies to death, then it must also apply to other supposed evils as well. Nagel gives a few different counter examples, and only mentions this one in passing, but it's a great example worthy of being expanded.

Suppose that you have two people, Bob and Dave. From their first person perspective, they can't tell the difference, but there is a difference. Bob is cheated on by their spouse, ridiculed by their friends, hated by their neighbors, and disparaged by their coworkers. Bob never is given a hint about this happening behind their back. Dave, on the other hand, has a faithful and loving spouse, their friends really do like them, their neighbors are sincere, and their coworkers actually respect them.

Who has the better life? Dave or Bob? One way to compare them is to ask yourself which life you would like to enter into, would you rather be reborn as Bob or Dave? If you think that Dave has the better life, then you are agreeing with Nagel that there are certain things which are bad for you despite you never actually experiencing them.

**When we are dead, we don't experience anything, there is no subject of the experience, so can we say that anything is good or bad for a person when they don't exist?**

Another way to put this is to say that, in the Bob vs Dave case, there is a subject, a person, who is actually missing out, but when a person is dead, there is no person to miss out on it, they don't exist. There are many things out there which require us to experience them in order for them to be either good or bad to us, for example, pleasure and pain. This, however, is not true for all cases. For example, a person who cheerfully pursues a life trying to communicate with plants would be wasting their time and we could call that a worse life than someone who cheerfully spends their life pursuing a vaccine for the common cold, even if neither actually succeed. The experience of the individuals in both of these cases is relatively the same, but there seems to be a difference in which life we would like to enter.

If we think that all goods and evils must be assigned to a person and experience, then we will have a hard time figuring out what makes things bad. Loss, betrayal, deception, and ridicule are on this view bad because people suffer when they learn of them. But it should be asked how our ideas of human value would have to be constituted to accommodate these cases directly instead. This view has the idea that we hurt when we learn of these things because they are bad, not that they are bad because we hurt when we learn of them.

**Why is there a difference between nonexistence prior to birth and nonexistence after birth?**

In this case, we say that the time we missed out on prior to our birth isn't good or bad (as in, we don't say that we missed out on that stuff) but the nonexistence after our death is something bad, according to Nagel. Some may think that this is a strange distinction. In both cases, we don't exist and there are experi-

ences which we could have had, so what is with the difference? Nagel has two different replies to this and they work together to refute this. First, Nagel contends that prior to a person's birth, if they are born at all, they aren't missing out on anything. This is because, with the exception of some premature births, a person born at a substantially different time, even with the same parents, would not be you. This is a stance which waxes and wanes in popularity called Origin Essentialism. Essentially, you could not have had different biological parents and could only come from that particular sperm and egg. Since that particular set up could have only happened in a particular way and at a particular time, there is very little you could have missed out on prior to your birth. The second part of this is a response to the Roman philosopher and poet Titus Lucretius Carus, better known as Lucretius. Lucretius makes the following claim, put into more ordinary words: Prior to your birth, you don't exist and after your death, you don't exist; so it's irrational to fear death just as much as it is irrational to fear the time prior to your birth. Lucretius, in this piece, is claiming that there is a temporal symmetry between these two times, so we should think of them the same. Nagel, however, thinks that it's asymmetrical, they aren't the same. The time after your death is time that you could have had, experiences which you could have experienced. You aren't missing out on the time prior to your birth but you are missing out on the time after your death.

## MODULE III

*What is The Mind?*

*What is the Body?*



# *The Mind-Body*

## *Problem by Tim Crane*

<sup>23</sup> The mind-body problem is the problem of explaining how our mental states, events and processes—like beliefs, actions and thinking—are related to the physical states, events and processes in our bodies. A question of the form, ‘how is A related to B?’ does not by itself pose a philosophical problem. To pose such a problem, there has to be something about A and B which makes the relation between them seem problematic. Many features of mind and body have been cited as responsible for our sense of the problem. Here I will concentrate on two: the fact that mind and body seem to interact causally, and the distinctive features of consciousness.

A long tradition in philosophy has held, with René Descartes, that the mind must be a non-bodily entity: a soul or mental substance. This thesis is called ‘substance dualism’ (or ‘Cartesian dualism’) because it says that there are two kinds of substance in the world, mental and physical or material. One reason for believing this is the belief that the soul, unlike the body, is immortal. Another reason for believing it is that we

---

<sup>23</sup>Tim Crane, “[The Mind-Body Problem](#),” *The MIT Encyclopedia of the Cognitive Sciences*, edited by Rob Wilson and Frank Keil (Cambridge, MA, USA: MIT P, 1999).

have free will, and this seems to require that the mind is a non-physical thing, since all physical things are subject to the laws of nature.

To say that the mind (or soul) is a mental substance is not to say that the mind is made up of some non-physical kind of stuff or material. The use of the term 'substance' is rather the traditional philosophical use: a substance is an entity which has properties and persists through change in its properties. A tiger, for instance, is a substance, whereas a hurricane is not. To say that there are mental substances— individual minds or souls—is to say that there are objects which are non-material or non-physical, and these objects can exist independently of physical objects, like a person's body. These objects, if they exist, are not made of non-physical 'stuff': they are not made of 'stuff' at all.

But if there are such objects, then how do they interact with physical objects? Our thoughts and other mental states often seem to be caused by events in the world external to our minds, and our thoughts and intentions seem to make our bodies move. A perception of a glass of wine can be caused by the presence of a glass of wine in front of me, and my desire for some wine plus the belief that there is a glass of wine in front of me can cause me to reach towards the glass. But many think that all physical effects are brought about by purely physical causes: physical states of my brain are enough to cause the physical event of my reaching towards the glass. So how can my mental states play any causal role in bringing about my actions?

Some dualists react to this by denying that such psychophysical causation really exists (this view is called 'epiphenomenalism'). Some philosophers have thought that mental states are causally related only to other mental states, and physical states are causally related only to other physical states: the mental and physical realms operate independently. This 'parallelism' view has been unpopular in the 20th century, as have most dualist views. For if we find dualism unsatisfactory, there is another way to answer the question of psychophysical causation:

we can say that mental states have effects in the physical world precisely because they are, contrary to appearances, physical states.<sup>24</sup> This is a \*monist\* view, since it holds that there is \*one\* kind of substance, physical or material substance. Therefore it is known as ‘physicalism’ or ‘materialism’.

Physicalism comes in many forms. The strongest form is the form just mentioned, which holds that mental \*states\* or \*properties\* are identical with physical states or properties. This view, sometimes called the ‘type-identity theory’, is considered an empirical hypothesis, awaiting confirmation by science. The model for such an identity theory is the identification of properties such as the heat of a gas with the mean kinetic energy of its constituent molecules. Since such an identification is often described as part of the \*reduction\* of thermodynamics to statistical mechanics, the parallel claim about the mental is often called a ‘reductive’ theory of mind, or ‘reductive physicalism’.<sup>25</sup>

Many philosophers find reductive physicalism an excessively bold empirical speculation. For it seems committed to the implausible claim that all creatures who believe that grass is green have one physical property in common—the property which is identical to the belief that grass is green. For this reason (and others) some physicalists adopt a weaker version of physicalism which does not have this consequence. This version of physicalism holds that all particular objects and events are physical, but allows that there are mental properties which are not identical to physical properties. (Davidson<sup>26</sup> is one inspiration for such views.) This kind of view, ‘non-reductive physicalism’, is a kind of dualism, since it holds there are two kinds of property, mental and physical. But it is not \*substance\* dualism, since it holds that all substances are physical substances.

---

<sup>24</sup>David Lewis, “An argument for the identity theory.” *Journal of Philosophy* vol. 63, 1966, pp. 17–25.

<sup>25</sup>David Lewis, “Reduction of mind,” *A Companion to the Philosophy of Mind*, edited by S. Guttenplan (Blackwell, 1995) pp. 412–31.

<sup>26</sup>Donald Davidson, “Mental Events,” *Experience and Theory*, edited by L. Foster and J. Swanson (Duckworth, 1970) pp. 79–101.

Non-reductive physicalism is also sometimes called a 'token-identity theory' since it identifies mental and physical particulars or tokens, and it is invariably supplemented by the claim that mental properties \*supervene\* on physical properties. Though the notion can be refined in many ways, supervenience is essentially a claim about the dependence of the mental on the physical: there can be no difference in mental facts without a difference in some physical facts (see Kim;<sup>27</sup> Horgan<sup>28</sup>).

If the problem of psychophysical causation was the whole of the mind-body problem, then it might seem that physicalism is a straightforward solution to that problem. If the only question is, 'how do mental states have effects in the physical world?', then it seems that the physicalist can answer this by saying that mental states are identical with physical states.

But there is a complication here. For it seems that physicalists can only propose this solution to the problem of psychophysical causation if mental causes are identical with physical causes. Yet if properties or states are causes, as many reductive physicalists assume, then non-reductive physicalists are not entitled to this solution, since they do not identify mental and physical properties. This is the problem of mental causation for non-reductive physicalists. (See Davidson,<sup>29</sup> Crane,<sup>30</sup> Jackson<sup>31</sup>).

However, even if the physicalist can solve this problem of mental causation, there is a deeper reason why there is more to the mind-body problem than the problem of psychophysical interaction. The reason is that, according to many philosophers, physicalism is not the \*solution\* to the mind-body problem, but

---

<sup>27</sup>J Kim, *Supervenience and Mind* (Cambridge UP, 1993).

<sup>28</sup>T. Horgan, "From supervenience to superdupervenience: meeting the demands of a material world," *Mind* vol. 102, 1993, pp. 555–86.

<sup>29</sup>Donald Davidson, "Thinking causes," *Mental Causation*, edited by J. Heil and A. Mele (Oxford UP, 1993) pp. 1–17.

<sup>30</sup>T. Crane, "The mental causation debate." *Proceedings of the Aristotelian Society, Supplementary Volume* vol. 69, 1995, pp. 211–36.

<sup>31</sup>F. Jackson, "Mental Causation," *Mind* vol. 105, 1996, pp. 377–413.

something which gives rise to a version of that problem. They reason as follows: we know enough to know that the world is completely physical. So if the mind exists, it too must be physical. However, it seems hard to understand how certain aspects of mind—notably consciousness—could just be physical features of the brain. How can the complex subjectivity of a conscious experience be produced by the grey matter of the brain? As McGinn<sup>32</sup> puts it, neurones and synapses seem ‘the wrong kind’ of material to produce consciousness. The problem here is one of intelligibility: we know that the mental is physical, so consciousness must have its origins in the brain; but how can we make sense of this mysterious fact?

Thomas Nagel dramatised this in a famous paper.<sup>33</sup> Nagel says that when a creature is conscious, there is something it is \*like\* to be that creature: there is something it is like to be a bat, but there is nothing it is like to be a stone. The heart of the mind-body problem for Nagel is the apparent fact that we cannot understand \*how\* consciousness can just be a physical property of the brain, even though we know that in some sense physicalism is true (see also Chalmers<sup>34</sup>).

Some physicalists respond by saying that this problem is illusory: if physicalism \*is\* true, then consciousness is just a physical property, and it simply begs the question against physicalism to wonder whether this \*can\* be true (see Lewis<sup>35</sup>). But Nagel’s criticism can be sharpened, as it has been by what Frank Jackson calls the ‘knowledge argument’ (Jackson;<sup>36</sup> see

---

<sup>32</sup>C. McGinn, “Can we solve the mind-body problem?” *Mind* vol. 98, 1989, pp. 349–66.

<sup>33</sup>T. Nagel, “What is it like to be a bat?” *Philosophical Review* vol. 4, 1974, pp. 435–50.

<sup>34</sup>D. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford UP, 1996).

<sup>35</sup>David Lewis, “Mad pain and martian pain,” *Philosophical Papers Volume One*, edited by D. Lewis (Oxford UP, 1983) pp. 122–32.

<sup>36</sup>F. Jackson, “Epiphenomenal qualia,” *Philosophical Quarterly* vol. 32, 1982, pp. 127–36.

also Robinson<sup>37</sup>). Jackson argues that even if we knew all the physical facts about, say, pain, we would not ipso facto know what it is like to be in pain. Someone omniscient about the physical facts about pain would learn something new when they learn what it is like to be in pain.

Therefore there is some knowledge—knowledge of what it is like—which is not knowledge of any physical fact. So not all facts are physical facts. (For physicalist responses to Jackson's argument see Lewis;<sup>38</sup> Dennett;<sup>39</sup> Churchland.<sup>40</sup>)

In late twentieth century philosophy of mind, discussions of the mind-body problem revolve around the twin poles of the problem of psychophysical causation and the problem of consciousness. And while it is possible to see these as independent problems, there is nonetheless a link between them, which can be expressed as a dilemma: if the mental is not physical, then how we make sense of its causal interaction with the physical? But if it is physical, how can we make sense of the phenomena of consciousness? These two questions, in effect, define the contemporary debate on the mind-body problem.

---

<sup>37</sup>H. Robinson, *Matter and Sense* (Cambridge UP, 1982).

<sup>38</sup>David Lewis, "What experience teaches," *Mind and Cognition*, edited by W. Lycan (Blackwell, 1990) pp. 499–519.

<sup>39</sup>D.C. Dennett, *Consciousness Explained* (Harmondsworth: Allen Lane, 1991).

<sup>40</sup>P.M. Churchland, "Reduction, qualia and the direct introspection of brain states," *Journal of Philosophy* vol. 82, 1985, pp. 8–28.

# *Part 4: The Mind-Body Problem*

The mind-body problem is a very, very, old puzzle in philosophy. It is trying to explain how our 'what-it's-like-nesses' relate to our bodily states. It's trying to explain how our thoughts, emotions, and other mental qualities relate to physical (bodily) events. Basically, the Mind-Body Problem is trying to answer a question of the form:

How is X related to Y?

There are many questions out there which are of this general form, for example, we have questions like:

How is the cause related to the effect?

How is exposition related to multiplication?

How is matter related to gravity?

How are the results of actions related to morality?

For the Mind-Body Problem, the question is:

How is the mind related to the body? How is the body related to the mind?

Answering the question “how does one thing relate to another?” isn’t too hard to answer in most contexts. To make it a problem worth thinking about, the things need to have some obvious connection between them *and* there needs to be something about the connection which seems problematic, or leads to problems. Conspiracy theories often try to make a connection between two things, but either there’s not an obvious connection or the proposed connection is problematic (crazy jumps in reasoning, odd consequences, etc.).<sup>41</sup>

In this case, the question “how is the mind related to the body?” is not easy to answer and just about any claimed relationship between the two seems to have problems. But, it seems very obvious that the mind, my ‘thinker’ so to speak, does have some kind of relationship with my body. I think of things, my body does them, I feel sad, my body cries. So, the core question is:

How are our mental states, beliefs, feelings, or thinkings, related to our bodily states, the events which go on physically?

There are two dominate theories, **physicalism** and **substance dualism**, about this and both have their problems, which we will go into detail about.

## Part 4.1: Substance Dualism

This is a very classic stance that there are two sorts of things in the world, rather than just one. It was best put forth by a guy named Rene Descartes (who we will go in depth on in Module 6). Some people think that there’s only material, physical, things in the world. Descartes thought that there were two kinds of things:

---

<sup>41</sup>If you are interested in the Philosophy of Conspiracy Theories, I have a collection of readings, and an order to read them in, which can be at an intro-level.



Physical Substances

Mental Substances

## Substance:

In that tradition of philosophy, a **substance** is a thing which has properties and can survive the change in those properties. For example, my car has the property 'silver' but I can paint it black and it would still be my car.

To say that there are mental substances is to say that there are non-material entities which exist independently of the physical stuff, like the body. The everyday term for a mental substance is 'soul'. We should be careful, this is a very confusing area; mental substances, or souls, are not composite according to the dualist. They are, to use some modern terminology, simple. Your typical physical substance is composed, built-up of, smaller physical things, like wood and metal, then the wood and metal are built of smaller parts, all the way down to strings (if String Theory is correct) or some other basic building-block. Mental substances (souls) aren't supposed to be like that. They are not composed of non-physical stuff and they are not composed of physical stuff, they are simple, basic, non-composite. This is to say that mental substances do not have parts, they cannot be broken up, so to speak, or divided.

**Substance dualism** is sort of the default view which many people have when they enter into this kind of debate, they think that they have a soul or something like that, and there have been various reasons given to think that there is this sort of soul.<sup>42</sup> Some people think that there is some sort of afterlife. Though it is possible to get an afterlife (with consciousness) without a soul, the default view seems to be that one is required. Others claim that people have free will and this requires a non-material soul. Others still think that there are different properties had by *you* and your *body*. If two things are

---

<sup>42</sup>This is especially true if the person enters the debate with a prior belief in an afterlife or certain religious stances.

the same, then there wouldn't be that kind of difference. For example:

- 1 I can imagine myself without a body.
- 2 If two things are the same, then if you imagine one, you imagine the other.
- 3 Therefore, I am not my body.

From this we get that I am a soul, something non-physical.

Another argument which is given for this stance goes like this:

- 1 The mind is immortal, the body is not.
- 2 If the mind and the body were the same thing, this would not be the case.
- 3 Therefore, the soul is different from the body.

And a third argument often given for this stance, again from Descartes, goes like this:

- 1 I have free will.
- 2 If I was just physical, then I would be subject to the laws of nature (no free will).
- 3 Therefore, I am not just physical.

This particular argument is the center of a current project I am working on concerning free-will and this problem. We will explore more of this (free will) in Module 4.

## **The Mary's Room Thought Experiment**

This is a relatively recent thought-experiment, a case to think about and try to understand with interesting implications. It was written in 1982 by Frank Jackson. The point of it is to give reason to think that everything is not just physical. This thought experiment leads to an argument for some flavor of dualism. I know that it is not very realistic, but I have amended it to make it more so. Here is the case:

Two undercover spies fell in love and had a child. Due to the nature of their work, the government took the child and locked her away in a room. The child is named “Mary”. Mary was forced to grow up in this room, but here’s the kicker, there’s only black and white. Mary never experiences colors at all. Mary, in this room, grows to be a brilliant scientist. She specializes in the neurophysiology of vision and acquires all the physical information there is to obtain about what goes on when we see a red rose, or the sky, and use terms like ‘red’, ‘blue’, and so on. She discovers, for example, just which wavelength combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence ‘The sky is blue’... Over the years, the political climate has changed and this sort of undercover work and the cloistering of the children becomes very taboo. The president finds out about Mary’s plight and orders her to be released. What will happen when Mary is released from her black and white room? On the day of her release, the president is present and hands her a red rose. Does she learn anything in that moment?

### **The Mary’s Room Argument**

Mary knows all of the physical facts about color vision.

Mary has never experienced color.

Upon seeing color for the first time, Mary learns something.

If you learn something, then that thing is a fact you did not know before.

So, Mary did not know a fact about color vision.

If Mary did not know a fact about color vision, that fact must be non-physical.

Therefore, there are some non-physical (mental) facts.

This argument will *either* get you substance dualism or a particular version of physicalism, but it will not get you certain kinds of physicalism. There are ways out for the physicalist, which we will see later.

### How Do These Substances Interact?

Now we need to ask how mental substances cause physical events and vice versa. The stance that there is this sort of causation between them is called **epiphenomenalism**. It seems clear that certain bodily states, like stubbing my toe, result in mental states (the feeling of pain) and that certain mental states, like feeling sad, result in bodily states (crying). But, how does this work?

The Causation Problem is essentially asking "how do body states cause mental states?" and, of even more importance, "how do mental states cause bodily states?" Many philosophers go with Epiphenomenalism because they think that we have some kind of free will. The various notions of free will and the arguments for and against it will come in Module 4. But, for now, let's just say that by 'free will' I mean that (at least some of) our choices are not deterministic, that some super-computer could not predict the choices we make before we make them.

If our choices are not deterministic, then they must be from outside of the laws of nature (because the laws of nature are deterministic (if they aren't, that doesn't help either)). If they come from outside of the laws of nature, then they must come from something non-physical (because the laws of nature govern physical things). So, something non-physical must be able to interact with something physical (EG the mental with the physical).

Boiling all of that last paragraph down, if our actions are non-deterministic, then something non-physical must be able to interact with something physical. This is far from a settled claim, as we will see when we cover free will. But, if you go with substance dualism to get free will, then you have a major

problem. You need to have a way of getting the mental substances and the physical substances to interact, at least mental to physical.

Some claim that this is an impossible task. The physical things are physical, the mental things are mental, and never the two shall meet. These folks claim that there can be no causation or interaction between the mental and physical, so by reasoning above, our actions are deterministic (uh-oh). Others claim that there can only be causation going from physical to mental, but not the other way around. This gives us the same problem as before.

### **Pre-Established Harmony**

There is one theory which accepts, whole hog, the idea that there are two kinds of things in the world, mental and physical, and that the two cannot affect each other. This view, put forth by Leibniz,<sup>43</sup> states that a person has two things, a soul and a physical body (a person is composed of a mental substance and a physical substance), but makes three further claims:

No state of a mind can cause a state in another mind or body and no body can cause a state in another body or mind (basically, minds and bodies can't interact, minds and minds can't interact, and bodies and bodies can't interact).

Every state of a substance which wasn't a miracle and wasn't its starting state, was caused by the previous state of that substance (basically, how some substance was determines how it will be).

Minds and bodies are programmed (or pre-determined) to behave in mutual coordination with each-other.

---

<sup>43</sup>references and citation needed

This is the Pre-Established Harmony stance. The first claim gives us our answer to the Causation Problem, namely, there's not any causation between the mental and the physical. Rather, because of the third claim, it merely appears to be causation. There's correlation but not causation. The second claim smooths out any wrinkles which may appear in the stance because it gives us that the world is deterministic, lights and clockwork.

For most versions of Pre-Established Harmony out there, it would seem, the 'programming' of the substances is arranged by God or some other divine architect.<sup>44</sup> Many people think that God is all-knowing, and this will appear again when we discuss arguments for and against the existence of God. If this is correct, then we can explain this by saying that God was the one who programmed the substances. But this leads to a further worry, and a potential problem.

As I mentioned before, many people like Substance Dualism because it will get you some kind of Free Will, but Pre-Established Harmony denies the possibility of a substance doing other than how it was programmed, everything in the world is deterministic. This denies the possibility of Free-Will. People will have the illusion of control, but that control is much like a little kid holding a toy steering wheel. Sure they may mimic, without realizing, the movements of the driver perfectly, but they are not the one driving the car.

## Part 4.2: Monism/Physicalism

Another way to solve the causation problem and Mind-Body Problem is to say that there's actually only one kind of substance in the world. This is a rejection of dualism and is called **monism**. It can come in two forms.

The first of these forms is called **Idealism**. This is the stance that there are only mental substances in the world. So, and I

---

<sup>44</sup>This can be related to divine foreknowledge as a problem for free will.

mean this jokingly, be nice to your table, it has feelings. The second is the real one which we will be concerned with for this class, but idealism does still have its adherents, is called **Physicalism**.

As a stance in the Mind-Body Problem as well as in other areas (though not as common), Physicalism comes in several different forms. But to really understand the distinction between these, we should cover what is meant by the terms "type" and "token." A type is a general class of things. For example, 'tree' is a type, same with 'car'. There are many individual things which are labeled as trees or as cars. Tokens, on the other hand, are individual instances of a type. When we talk about various things, it's useful to be careful about whether we are dealing with types or with tokens. For example, if someone were to claim that 'lying is morally wrong', we would need to know whether they are talking about all cases where a person knowingly misinforms another or just an individual instance of doing so. If they are talking about type of action and labeling all cases of lying as wrong, then all we need to do is point to a cases where it's OK to lie and that would disprove their claim. On the other hand, if they are talking about a token of the action, then we would need to look closely at the individual case to try and prove them wrong.

## Reductive Physicalism

The different kinds of physicalism all share something in common, namely, that the mental states are the physical states, but they differ in whether they think in terms of types or tokens. The first, called "Reductive Physicalism" thinks in terms of types, they claim that the mental states you have are identical to your physical states, meaning that a type of mental experience maps to a type of physical state (likely in the brain). This kind of theory is far more empirical in nature than the others which philosophers typically deal with and would require a ton of brain scans to set up. The model for this kind of identity

(type-identity) is often found in science, where they identify a general class of things with another. For example, how hot an object is and the mean kinetic energy of its molecules.

Reductive Physicalism is the strongest one which you are going to find, it's making a very bold claim. Some claim that this is too bold of a claim. Reductive Physicalism entails that if two people have the same thought, then their brains had to be lit up (so to speak) in the same way. To some, this just doesn't seem plausible, as people come to the same conclusion about things all the time, but all brains are fundamentally different. One could bite the bullet and say that people have similar thoughts, but never the same thought, but that too requires a ton of experimental data.

All that being said, the Reductive Physicalist does have a solution to the Mind Body Problem and the Causation Problem. For the Mind-Body Problem, the solution is that the mind is the body, there's no difference and for the Causation Problem, it's just physical to physical causation, so who cares?

## **Non-Reductive Physicalism**

Some people want to keep the physicalism, but don't want to say that all people have different thoughts, that two people can never have the same thought. This is where we get the other major kind of physicalism, Non-Reductive Physicalism. Rather than dealing with types, Non-Reductive Physicalism deals with tokens. Like Reductive Physicalism, Non-Reductive Physicalism says that there's only one kind of substance, namely physical, but Non-Reductive Physicalism claims that there are two different kinds of properties.

One way to think about this is in terms of colors. There are many different ways in which a certain shade of green can be produced. This can be from the particles on the surface of an object being arranged a certain way and having uncolored light bounce off of it or it can be from having colored light bouncing off of it and its particles being arranged in a different way.



But, regardless, it's still the same shade of green. Similarly, the mental state, your thought, can be produced from a whole bunch of different arrangements of neurons in your brain. For the Non-Reductive Physicalist, the identification between the mind and the body is one of supervenience.

Supervenience is a bit of a tricky topic, but mostly because it's a word that you hardly ever see, we encounter the concept all the time without realizing it. Supervenience is a kind of relation between two things. It is basically that one thing supervenes on another when there can't be a change in the first without a change in the second. The first depends on the second. So, for example, whether or not something is beautiful supervenes on its arrangement. If you want to make something more or less beautiful, you fiddle with how it's arranged. Similarly, some claim that the morality of an action supervenes on the results, so if you want to make the right action, choose the one with the best results, you can't change the morality of an action without changing what the results of it were. Looking a little more politically, societies supervene on the people. So, if you want to change a society, you need to change the people in it (typically this means convince them of something). And, finally, if the color example worked for you, the color of an object supervenes on the arrangement of the particles on the surface and the light striking it.

Going back to the point at hand, Non-Reductive Physicalism claims that the mental supervenes on the physical, meaning that there can be no change in the mental without a change in the physical. If you want to see this in action, look at videos of people getting fMRI scans. This is a sort of have your cake and eat it too kind of stance, they get that there is something mental 'up-stairs', but they also get all of the scientific power of Physicalism. This is likely the reason why most philosophers today are Non-Reductive Physicalists.

Non-Reductive Physicalism gets all of the same answers to the Mind-Body Problem as well as the Causation Problem as the Reductive Physicalist, but it is less committed to such bold

claims and it gets various other fun results.

## Current Developments

Non-Reductive Physicalism and Reductive Physicalism both have the massive support which they do for a reason, but they don't paint the whole picture. If the question of the Mind-Body Problem was just 'how does the mind interact with the body (and vice versa)?' then the Physicalist, of either form, has an easy answer and would seem right. But assuming Physicalism is right, this leads to another problem.

We know enough to know that the world is completely physical. So if the mind exists, it too must be physical. However, it seems hard to understand how certain aspects of mind—notably consciousness—could just be physical features of the brain. How can the complex subjectivity of a conscious experience be produced by the grey matter of the brain?

This is the biggest question in the Philosophy of Mind at the moment, I know it as the Hard Problem of Consciousness, basically, it's asking 'how do physical things make something as complicated as the mind?' Assuming that everything is physical, where does consciousness come from? This is the end of the road, paths here are still being paved. In the next section of this module, we will explore this problem, and others, in relation to contemporary computer science and artificial intelligence.

There is also a stance, which is an epistemological one, related to this which is called mysterianism which states that the problem of consciousness and the Mind-Body Problem in general is not possible for us to solve. It can also be phrased as saying that we know enough about the world to know that Physicalism is true, but how/why this is the case is beyond the ability of the human brain/mind to answer. Remember that an

epistemological question is one which concerns whether or not one can or does know something where as a metaphysical question is about whether or not something is the case. Mysterians hold that Physicalism is true, which is a metaphysical stance, but at the same time say that it is impossible for us to know how that works.

# *Minds, Brains and Programs by John Searle*

45

## **Abstract**

This article can be viewed as an attempt to explore the consequences of two propositions. (1) Intentionality in human beings (and animals) is a product of causal features of the brain I assume this is an empirical fact about the actual causal relations between mental processes and brains It says simply that certain brain processes are sufficient for intentionality. (2) Instantiating a computer program is never by itself a sufficient condition of intentionality The main argument of this paper is directed at establishing this claim The form of the argument is to show how a human agent could instantiate the program and still not have the relevant intentionality. These two propositions have

---

<sup>45</sup>John Searle, “Minds, Brains, and Programs,” *Behavioral and Brain Sciences*, 1980, pp. 417–57.

the following consequences (3) The explanation of how the brain produces intentionality cannot be that it does it by instantiating a computer program. This is a strict logical consequence of 1 and 2. (4) Any mechanism capable of producing intentionality must have causal powers equal to those of the brain. This is meant to be a trivial consequence of 1. (5) Any attempt literally to create intentionality artificially (strong AI) could not succeed just by designing programs but would have to duplicate the causal powers of the human brain. This follows from 2 and 4.

"Could a machine think?" On the argument advanced here only a machine could think, and only very special kinds of machines, namely brains and machines with internal causal powers equivalent to those of brains. And that is why strong AI has little to tell us about thinking, since it is not about machines but about programs, and no program by itself is sufficient for thinking.

## Minds, Brains, and Programs

What psychological and philosophical significance should we attach to recent efforts at computer simulations of human cognitive capacities? In answering this question, I find it useful to distinguish what I will call "strong" AI from "weak" or "cautious" AI (Artificial Intelligence). According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explana-

tions; rather, the programs are themselves the explanations.

I have no objection to the claims of weak AI, at least as far as this article is concerned. My discussion here will be directed at the claims I have defined as those of strong AI, specifically the claim that the appropriately programmed computer literally has cognitive states and that the programs thereby explain human cognition. When I hereafter refer to AI, I have in mind the strong version, as expressed by these two claims.

I will consider the work of Roger Schank and his colleagues at Yale,<sup>46</sup> because I am more familiar with it than I am with any other similar claims, and because it provides a very clear example of the sort of work I wish to examine. But nothing that follows depends upon the details of Schank's programs. The same arguments would apply to Winograd's SHRDLU,<sup>47</sup> Weizenbaum's ELIZA,<sup>48</sup> and indeed any Turing machine simulation of human mental phenomena.

Very briefly, and leaving out the various details, one can describe Schank's program as follows: the aim of the program is to simulate the human ability to understand stories. It is characteristic of human beings' story- understanding capacity that they can answer questions about the story even though the information that they give was never explicitly stated in the story. Thus, for example, suppose you are given the following story:

-A man went into a restaurant and ordered a hamburger. When the hamburger arrived it was burned to a crisp, and the man stormed out of the restaurant angrily, without paying for the hamburger or leaving a tip." Now, if you are asked -

---

<sup>46</sup>R. C. Schank and R. P. Abelson, *Scripts, plans, goals, and understanding* (Erlbaum P, 1977).

<sup>47</sup>T. Winograd, "A procedural model of language understanding," *Computer models of thought and language*, edited by R. Schank and K. Colby (Freeman, 1973) pp. 152-89.

<sup>48</sup>J. Weizenbaum, "Eliza - a computer program for the study of natural language communication between man and machine," *Communication of the Association for Computing Machinery* vol. 9, 1965, pp. 36-45.

Did the man eat the hamburger?" you will presumably answer, 'No, he did not.' Similarly, if you are given the following story: '-A man went into a restaurant and ordered a hamburger; when the hamburger came he was very pleased with it; and as he left the restaurant he gave the waitress a large tip before paying his bill,' and you are asked the question, '-Did the man eat the hamburger?,-' you will presumably answer, '-Yes, he ate the hamburger.'" Now Schank's machines can similarly answer questions about restaurants in this fashion. To do this, they have a "-representation" of the sort of information that human beings have about restaurants, which enables them to answer such questions as those above, given these sorts of stories. When the machine is given the story and then asked the question, the machine will print out answers of the sort that we would expect human beings to give if told similar stories. Partisans of strong AI claim that in this question and answer sequence the machine is not only simulating a human ability but also

1. that the machine can literally be said to understand the story and provide the answers to questions, and
2. that what the machine and its program do explains the human ability to understand the story and answer questions about it.

Both claims seem to me to be totally unsupported by Schank's work, as I will attempt to show in what follows. One way to test any theory of the mind is to ask oneself what it would be like if my mind actually worked on the principles that the theory says all minds work on. Let us apply this test to the Schank program with the following Gedankenexperiment. Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and that I'm not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles. To me, Chinese writing is just so many meaningless squiggles.

Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal symbols, and all that 'formal' means here is that I can identify the symbols entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch. Unknown to me, the people who are giving me all of these symbols call the first batch "a script," they call the second batch a "story. ' and they call the third batch "questions." Furthermore, they call the symbols I give them back in response to the third batch "answers to the questions." and the set of rules in English that they gave me, they call "the program."

Now just to complicate the story a little, imagine that these people also give me stories in English, which I understand, and they then ask me questions in English about these stories, and I give them back answers in English. Suppose also that after a while I get so good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view that is, from the point of view of somebody outside the room in which I am locked – my answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at my answers can tell that I don't speak a word of Chinese.

Let us also suppose that my answers to the English questions are, as they no doubt would be, indistinguishable from those of other native English speakers, for the simple reason that I am a native English speaker. From the external point



of view – from the point of view of someone reading my "answers" – the answers to the Chinese questions and the English questions are equally good. But in the Chinese case, unlike the English case, I produce the answers by manipulating uninterpreted formal symbols. As far as the Chinese is concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purposes of the Chinese, I am simply an instantiation of the computer program.

Now the claims made by strong AI are that the programmed computer understands the stories and that the program in some sense explains human understanding. But we are now in a position to examine these claims in light of our thought experiment.

1 As regards the first claim, it seems to me quite obvious in the example that I do not understand a word of the Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, and I can have any formal program you like, but I still understand nothing. For the same reasons, Schank's computer understands nothing of any stories. whether in Chinese. English. or whatever. since in the Chinese case the computer is me. and in cases where the computer is not me, the computer has nothing more than I have in the case where I understand nothing.

2. As regards the second claim, that the program explains human understanding, we can see that the computer and its program do not provide sufficient conditions of understanding since the computer and the program are functioning, and there is no understanding. But does it even provide a necessary condition or a significant contribution to understanding? One of the claims made by the supporters of strong AI is that when I understand a story in English, what I am doing is exactly the same – or perhaps more of the same – as what I was doing in manipulating the Chinese symbols. It is simply more formal symbol manipulation that distinguishes the case in English, where I do understand, from the case in Chinese, where I don't. I have not demonstrated that this claim is false, but it would certainly appear an incredible claim in the example.

Such plausibility as the claim has derives from the supposition that we can construct a program that will have the same inputs and outputs as native speakers, and in addition we assume that speakers have some level of description where they are also instantiations of a program.

On the basis of these two assumptions we assume that even if Schank's program isn't the whole story about understanding, it may be part of the story. Well, I suppose that is an empirical possibility, but not the slightest reason has so far been given to believe that it is true, since what is suggested though certainly not demonstrated – by the example is that the computer program is simply irrelevant to my understanding of the story. In the Chinese case I have everything that artificial intelligence can put into me by way of a program, and I understand nothing; in the English case I understand everything, and there is so far no reason at all to suppose that my understanding has anything to do with computer programs, that is, with computational operations on purely formally specified elements. As long as the program is defined in terms of computational operations on purely formally defined elements, what the example suggests is that these by themselves have no interesting connection with understanding. They are certainly not sufficient conditions, and not the slightest reason has been given to suppose that they are necessary conditions or even that they make a significant contribution to understanding.

Notice that the force of the argument is not simply that different machines can have the same input and output while operating on different formal principles – that is not the point at all. Rather, whatever purely formal principles you put into the computer, they will not be sufficient for understanding, since a human will be able to follow the formal principles without understanding anything. No reason whatever has been offered to suppose that such principles are necessary or even contributory, since no reason has been given to suppose that when I understand English I am operating with any formal program at all.

Well, then, what is it that I have in the case of the English sentences that I do not have in the case of the Chinese sentences? The obvious answer is that I know what the former mean, while I haven't the faintest idea what the latter mean. But in what does this consist and why couldn't we give it to a machine, whatever it is? I will return to this question later, but first I want to continue with the example.

I have had the occasions to present this example to several workers in artificial intelligence, and, interestingly, they do not seem to agree on what the proper reply to it is. I get a surprising variety of replies, and in what follows I will consider the most common of these (specified along with their geographic origins).

But first I want to block some common misunderstandings about "understanding": in many of these discussions one finds a lot of fancy footwork about the word "understanding." My critics point out that there are many different degrees of understanding; that "understanding" is not a simple two-place predicate; that there are even different kinds and levels of understanding, and often the law of excluded middle doesn't even apply in a straightforward way to statements of the form "x understands y; that in many cases it is a matter for decision and not a simple matter of fact whether x understands y; and so on. To all of these points I want to say: of course, of course. But they have nothing to do with the points at issue. There are clear cases in which "understanding" literally applies and clear cases in which it does not apply; and these two sorts of cases are all I need for this argument. I understand stories in English; to a lesser degree I can understand stories in French; to a still lesser degree, stories in German; and in Chinese, not at all. My car and my adding machine, on the other hand, understand nothing: they are not in that line of business. We often attribute "understanding" and other cognitive predicates by metaphor and analogy to cars, adding machines, and other artifacts, but nothing is proved by such attributions. We say, "The door knows when to open because of its photoelectric cell," "The adding machine knows how) (understands how to,

is able) to do addition and subtraction but not division,” and “The thermostat perceives changes in the temperature.”

The reason we make these attributions is quite interesting, and it has to do with the fact that in artifacts we extend our own intentionality;<sup>3</sup> our tools are extensions of our purposes, and so we find it natural to make metaphorical attributions of intentionality to them; but I take it no philosophical ice is cut by such examples. The sense in which an automatic door “understands instructions” from its photoelectric cell is not at all the sense in which I understand English. If the sense in which Schank’s programmed computers understand stories is supposed to be the metaphorical sense in which the door understands, and not the sense in which I understand English, the issue would not be worth discussing. But Newell and Simon (1963) write that the kind of cognition they claim for computers is exactly the same as for human beings. I like the straightforwardness of this claim, and it is the sort of claim I will be considering. I will argue that in the literal sense the programmed computer understands what the car and the adding machine understand, namely, exactly nothing. The computer understanding is not just (like my understanding of German) partial or incomplete; it is zero. Now to the replies:

I. The systems reply (Berkeley). “While it is true that the individual person who is locked in the room does not understand the story, the fact is that he is merely part of a whole system, and the system does understand the story. The person has a large ledger in front of him in which are written the rules, he has a lot of scratch paper and pencils for doing calculations, he has ‘data banks’ of sets of Chinese symbols. Now, understanding is not being ascribed to the mere individual; rather it is being ascribed to this whole system of which he is a part.” My response to the systems theory is quite simple: let the individual internalize all of these elements of the system. He memorizes the rules in the ledger and the data banks of Chinese symbols, and he does all the calculations in his head. The individual then incorporates the entire system. There isn’t anything at all

to the system that he does not encompass. We can even get rid of the room and suppose he works outdoors. All the same, he understands nothing of the Chinese, and a fortiori neither does the system, because there isn't anything in the system that isn't in him. If he doesn't understand, then there is no way the system could understand because the system is just a part of him.

Actually I feel somewhat embarrassed to give even this answer to the systems theory because the theory seems to me so implausible to start with. The idea is that while a person doesn't understand Chinese, somehow the conjunction of that person and bits of paper might understand Chinese. It is not easy for me to imagine how someone who was not in the grip of an ideology would find the idea at all plausible. Still, I think many people who are committed to the ideology of strong AI will in the end be inclined to say something very much like this; so let us pursue it a bit further. According to one version of this view, while the man in the internalized systems example doesn't understand Chinese in the sense that a native Chinese speaker does (because, for example, he doesn't know that the story refers to restaurants and hamburgers, etc.), still "the man as a formal symbol manipulation system" really does understand Chinese. The subsystem of the man that is the formal symbol manipulation system for Chinese should not be confused with the subsystem for English.

So there are really two subsystems in the man; one understands English, the other Chinese, and "it's just that the two systems have little to do with each other." But, I want to reply, not only do they have little to do with each other, they are not even remotely alike. The subsystem that understands English (assuming we allow ourselves to talk in this jargon of "subsystems" for a moment) knows that the stories are about restaurants and eating hamburgers, he knows that he is being asked questions about restaurants and that he is answering questions as best he can by making various inferences from the content of the story, and so on. But the Chinese system knows none of

this. Whereas the English subsystem knows that "hamburgers" refers to hamburgers, the Chinese subsystem knows only that "squiggle squiggle" is followed by "squoggle squoggle." All he knows is that various formal symbols are being introduced at one end and manipulated according to rules written in English, and other symbols are going out at the other end.

The whole point of the original example was to argue that such symbol manipulation by itself couldn't be sufficient for understanding Chinese in any literal sense because the man could write "squoggle squoggle" after "squiggle squiggle" without understanding anything in Chinese. And it doesn't meet that argument to postulate subsystems within the man, because the subsystems are no better off than the man was in the first place; they still don't have anything even remotely like what the English-speaking man (or subsystem) has. Indeed, in the case as described, the Chinese subsystem is simply a part of the English subsystem, a part that engages in meaningless symbol manipulation according to rules in English.

Let us ask ourselves what is supposed to motivate the systems reply in the first place; that is, what independent grounds are there supposed to be for saying that the agent must have a subsystem within him that literally understands stories in Chinese? As far as I can tell the only grounds are that in the example I have the same input and output as native Chinese speakers and a program that goes from one to the other. But the whole point of the examples has been to try to show that that couldn't be sufficient for understanding, in the sense in which I understand stories in English, because a person, and hence the set of systems that go to make up a person, could have the right combination of input, output, and program and still not understand anything in the relevant literal sense in which I understand English.

The only motivation for saying there must be a subsystem in me that understands Chinese is that I have a program and I can pass the Turing test; I can fool native Chinese speakers. But precisely one of the points at issue is the adequacy of

the Turing test. The example shows that there could be two "systems," both of which pass the Turing test, but only one of which understands; and it is no argument against this point to say that since they both pass the Turing test they must both understand, since this claim fails to meet the argument that the system in me that understands English has a great deal more than the system that merely processes Chinese. In short, the systems reply simply begs the question by insisting without argument that the system must understand Chinese.

Furthermore, the systems reply would appear to lead to consequences that are independently absurd. If we are to conclude that there must be cognition in me on the grounds that I have a certain sort of input and output and a program in between, then it looks like all sorts of noncognitive subsystems are going to turn out to be cognitive. For example, there is a level of description at which my stomach does information processing, and it instantiates any number of computer programs, but I take it we do not want to say that it has any understanding.<sup>49</sup> But if we accept the systems reply, then it is hard to see how we avoid saying that stomach, heart, liver, and so on, are all understanding subsystems, since there is no principled way to distinguish the motivation for saying the Chinese subsystem understands from saying that the stomach understands. It is, by the way, not an answer to this point to say that the Chinese system has information as input and output and the stomach has food and food products as input and output, since from the point of view of the agent, from my point of view, there is no information in either the food or the Chinese – the Chinese is just so many meaningless squiggles. The information in the Chinese case is solely in the eyes of the programmers and the interpreters, and there is nothing to prevent them from treating the input and output of my digestive organs as information if they so desire.

---

<sup>49</sup>Z. W. Pylyshyn, "Computation and cognition: issues in the foundations of cognitive science," *Behavioral and Brain Sciences* vol. 3, 1980,

This last point bears on some independent problems in strong AI, and it is worth digressing for a moment to explain it. If strong AI is to be a branch of psychology, then it must be able to distinguish those systems that are genuinely mental from those that are not. It must be able to distinguish the principles on which the mind works from those on which nonmental systems work; otherwise it will offer us no explanations of what is specifically mental about the mental. And the mental-nonmental distinction cannot be just in the eye of the beholder but it must be intrinsic to the systems; otherwise it would be up to any beholder to treat people as nonmental and, for example, hurricanes as mental if he likes. But quite often in the AI literature the distinction is blurred in ways that would in the long run prove disastrous to the claim that AI is a cognitive inquiry. McCarthy, for example, writes, 'Machines as simple as thermostats can be said to have beliefs, and having beliefs seems to be a characteristic of most machines capable of problem solving performance'.<sup>50</sup>

Anyone who thinks strong AI has a chance as a theory of the mind ought to ponder the implications of that remark. We are asked to accept it as a discovery of strong AI that the hunk of metal on the wall that we use to regulate the temperature has beliefs in exactly the same sense that we, our spouses, and our children have beliefs, and furthermore that "most" of the other machines in the room – telephone, tape recorder, adding machine, electric light switch, – also have beliefs in this literal sense. It is not the aim of this article to argue against McCarthy's point, so I will simply assert the following without argument. The study of the mind starts with such facts as that humans have beliefs, while thermostats, telephones, and adding machines don't. If you get a theory that denies this point you have produced a counterexample to the theory and the theory

---

<sup>50</sup>John McCarthy, "Ascribing Mental Qualities to Machines," *Philosophical Perspectives in Artificial Intelligence*, edited by Martin Ringle (Humanities P, 1979).



is false.

One gets the impression that people in AI who write this sort of thing think they can get away with it because they don't really take it seriously, and they don't think anyone else will either. I propose for a moment at least, to take it seriously. Think hard for one minute about what would be necessary to establish that that hunk of metal on the wall over there had real beliefs beliefs with direction of fit, propositional content, and conditions of satisfaction; beliefs that had the possibility of being strong beliefs or weak beliefs; nervous, anxious, or secure beliefs; dogmatic, rational, or superstitious beliefs; blind faiths or hesitant cogitations; any kind of beliefs. The thermostat is not a candidate. Neither is stomach, liver adding machine, or telephone. However, since we are taking the idea seriously, notice that its truth would be fatal to strong AI's claim to be a science of the mind. For now the mind is everywhere. What we wanted to know is what distinguishes the mind from thermostats and livers. And if McCarthy were right, strong AI wouldn't have a hope of telling us that.

II. The Robot Reply (Yale). "Suppose we wrote a different kind of program from Schank's program. Suppose we put a computer inside a robot, and this computer would not just take in formal symbols as input and give out formal symbols as output, but rather would actually operate the robot in such a way that the robot does something very much like perceiving, walking, moving about, hammering nails, eating drinking – anything you like. The robot would, for example have a television camera attached to it that enabled it to 'see,' it would have arms and legs that enabled it to 'act,' and all of this would be controlled by its computer 'brain.' Such a robot would, unlike Schank's computer, have genuine understanding and other mental states."

The first thing to notice about the robot reply is that it tacitly concedes that cognition is not solely a matter of formal symbol manipulation, since this reply adds a set of causal re-

lation with the outside world.<sup>51</sup> But the answer to the robot reply is that the addition of such "perceptual" and "motor" capacities adds nothing by way of understanding, in particular, or intentionality, in general, to Schank's original program. To see this, notice that the same thought experiment applies to the robot case. Suppose that instead of the computer inside the robot, you put me inside the room and, as in the original Chinese case, you give me more Chinese symbols with more instructions in English for matching Chinese symbols to Chinese symbols and feeding back Chinese symbols to the outside. Suppose, unknown to me, some of the Chinese symbols that come to me come from a television camera attached to the robot and other Chinese symbols that I am giving out serve to make the motors inside the robot move the robot's legs or arms. It is important to emphasize that all I am doing is manipulating formal symbols: I know none of these other facts. I am receiving "information" from the robot's "perceptual" apparatus, and I am giving out "instructions" to its motor apparatus without knowing either of these facts. I am the robot's homunculus, but unlike the traditional homunculus, I don't know what's going on. I don't understand anything except the rules for symbol manipulation. Now in this case I want to say that the robot has no intentional states at all; it is simply moving about as a result of its electrical wiring and its program. And furthermore, by instantiating the program I have no intentional states of the relevant type. All I do is follow formal instructions about manipulating formal symbols.

III. The brain simulator reply (Berkeley and M.I.T.). "Suppose we design a program that doesn't represent information that we have about the world, such as the information in Schank's scripts, but simulates the actual sequence of neuron firings at the synapses of the brain of a native Chinese speaker

---

<sup>51</sup>J. A. Fodor, "Methodological solipsism considered as a research strategy in cognitive psychology," *Behavioral and Brain Sciences* vol. 3, no. 1, 1980, <https://doi.org/10.1017/S0140525X00001771>, pp. 63–73.

when he understands stories in Chinese and gives answers to them. The machine takes in Chinese stories and questions about them as input, it simulates the formal structure of actual Chinese brains in processing these stories, and it gives out Chinese answers as outputs. We can even imagine that the machine operates, not with a single serial program, but with a whole set of programs operating in parallel, in the manner that actual human brains presumably operate when they process natural language. Now surely in such a case we would have to say that the machine understood the stories; and if we refuse to say that, wouldn't we also have to deny that native Chinese speakers understood the stories? At the level of the synapses, what would or could be different about the program of the computer and the program of the Chinese brain?"

Before countering this reply I want to digress to note that it is an odd reply for any partisan of artificial intelligence (or functionalism, etc.) to make: I thought the whole idea of strong AI is that we don't need to know how the brain works to know how the mind works. The basic hypothesis, or so I had supposed, was that there is a level of mental operations consisting of computational processes over formal elements that constitute the essence of the mental and can be realized in all sorts of different brain processes, in the same way that any computer program can be realized in different computer hardware: on the assumptions of strong AI, the mind is to the brain as the program is to the hardware, and thus we can understand the mind without doing neurophysiology. If we had to know how the brain worked to do AI, we wouldn't bother with AI. However, even getting this close to the operation of the brain is still not sufficient to produce understanding. To see this, imagine that instead of a mono lingual man in a room shuffling symbols we have the man operate an elaborate set of water pipes with valves connecting them. When the man receives the Chinese symbols, he looks up in the program, written in English, which valves he has to turn on and off. Each water connection corresponds to a synapse in the Chinese brain, and the whole system

is rigged up so that after doing all the right firings, that is after turning on all the right faucets, the Chinese answers pop out at the output end of the series of pipes.

Now where is the understanding in this system? It takes Chinese as input, it simulates the formal structure of the synapses of the Chinese brain, and it gives Chinese as output. But the man certainly doesn't understand Chinese, and neither do the water pipes, and if we are tempted to adopt what I think is the absurd view that somehow the conjunction of man and water pipes understands, remember that in principle the man can internalize the formal structure of the water pipes and do all the "neuron firings" in his imagination. The problem with the brain simulator is that it is simulating the wrong things about the brain. As long as it simulates only the formal structure of the sequence of neuron firings at the synapses, it won't have simulated what matters about the brain, namely its causal properties, its ability to produce intentional states. And that the formal properties are not sufficient for the causal properties is shown by the water pipe example: we can have all the formal properties carved off from the relevant neurobiological causal properties.

IV. The combination reply (Berkeley and Stanford). 'While each of the previous three replies might not be completely convincing by itself as a refutation of the Chinese room counterexample, if you take all three together they are collectively much more convincing and even decisive. Imagine a robot with a brain-shaped computer lodged in its cranial cavity, imagine the computer programmed with all the synapses of a human brain, imagine the whole behavior of the robot is indistinguishable from human behavior, and now think of the whole thing as a unified system and not just as a computer with inputs and outputs. Surely in such a case we would have to ascribe intentionality to the system.

I entirely agree that in such a case we would find it rational and indeed irresistible to accept the hypothesis that the robot had intentionality, as long as we knew nothing more about it.

Indeed, besides appearance and behavior, the other elements of the combination are really irrelevant. If we could build a robot whose behavior was indistinguishable over a large range from human behavior, we would attribute intentionality to it, pending some reason not to. We wouldn't need to know in advance that its computer brain was a formal analogue of the human brain.

But I really don't see that this is any help to the claims of strong AI; and here's why: According to strong AI, instantiating a formal program with the right input and output is a sufficient condition of, indeed is constitutive of, intentionality. As Newell<sup>52</sup> puts it, the essence of the mental is the operation of a physical symbol system. But the attributions of intentionality that we make to the robot in this example have nothing to do with formal programs. They are simply based on the assumption that if the robot looks and behaves sufficiently like us, then we would suppose, until proven otherwise, that it must have mental states like ours that cause and are expressed by its behavior and it must have an inner mechanism capable of producing such mental states. If we knew independently how to account for its behavior without such assumptions we would not attribute intentionality to it especially if we knew it had a formal program. And this is precisely the point of my earlier reply to objection 11.

Suppose we knew that the robot's behavior was entirely accounted for by the fact that a man inside it was receiving uninterpreted formal symbols from the robot's sensory receptors and sending out uninterpreted formal symbols to its motor mechanisms, and the man was doing this symbol manipulation in accordance with a bunch of rules. Furthermore, suppose the man knows none of these facts about the robot, all he knows

---

<sup>52</sup>Allen Newell, "Physical Symbol Systems\*," *Cognitive Science* vol. 4, no. 2. [https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0402\\_2](https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0402_2), 1980, [https://doi.org/https://doi.org/10.1207/s15516709cog0402\\_2](https://doi.org/https://doi.org/10.1207/s15516709cog0402_2), pp. 135–83.

is which operations to perform on which meaningless symbols. In such a case we would regard the robot as an ingenious mechanical dummy. The hypothesis that the dummy has a mind would now be unwarranted and unnecessary, for there is now no longer any reason to ascribe intentionality to the robot or to the system of which it is a part (except of course for the man's intentionality in manipulating the symbols). The formal symbol manipulations go on, the input and output are correctly matched, but the only real locus of intentionality is the man, and he doesn't know any of the relevant intentional states; he doesn't, for example, see what comes into the robot's eyes, he doesn't intend to move the robot's arm, and he doesn't understand any of the remarks made to or by the robot. Nor, for the reasons stated earlier, does the system of which man and robot are a part.

To see this point, contrast this case with cases in which we find it completely natural to ascribe intentionality to members of certain other primate species such as apes and monkeys and to domestic animals such as dogs. The reasons we find it natural are, roughly, two: we can't make sense of the animal's behavior without the ascription of intentionality and we can see that the beasts are made of similar stuff to ourselves – that is an eye, that a nose, this is its skin, and so on. Given the coherence of the animal's behavior and the assumption of the same causal stuff underlying it, we assume both that the animal must have mental states underlying its behavior, and that the mental states must be produced by mechanisms made out of the stuff that is like our stuff. We would certainly make similar assumptions about the robot unless we had some reason not to, but as soon as we knew that the behavior was the result of a formal program, and that the actual causal properties of the physical substance were irrelevant we would abandon the assumption of intentionality. [See "Cognition and Consciousness in Nonhuman Species BBS 1(4) 1978.]

There are two other responses to my example that come up frequently (and so are worth discussing) but really miss the

point.

V. The other minds reply (Yale). "How do you know that other people understand Chinese or anything else? Only by their behavior. Now the computer can pass the behavioral tests as well as they can (in principle), so if you are going to attribute cognition to other people you must in principle also attribute it to computers. ' This objection really is only worth a short reply. The problem in this discussion is not about how I know that other people have cognitive states, but rather what it is that I am attributing to them when I attribute cognitive states to them. The thrust of the argument is that it couldn't be just computational processes and their output because the computational processes and their output can exist without the cognitive state. It is no answer to this argument to feign anesthesia. In 'cognitive sciences" one presupposes the reality and knowability of the mental in the same way that in physical sciences one has to presuppose the reality and knowability of physical objects.

VI. The many mansions reply (Berkeley). "Your whole argument presupposes that AI is only about analogue and digital computers. But that just happens to be the present state of technology. Whatever these causal processes are that you say are essential for intentionality (assuming you are right), eventually we will be able to build devices that have these causal processes, and that will be artificial intelligence. So your arguments are in no way directed at the ability of artificial intelligence to produce and explain cognition." I really have no objection to this reply save to say that it in effect trivializes the project of strong AI by redefining it as whatever artificially produces and explains cognition. The interest of the original claim made on behalf of artificial intelligence is that it was a precise, well defined thesis: mental processes are computational processes over formally defined elements. I have been concerned to challenge that thesis. If the claim is redefined so that it is no longer that thesis, my objections no longer apply because there is no longer a testable hypothesis for them to apply to.

Let us now return to the question I promised I would try

to answer: granted that in my original example I understand the English and I do not understand the Chinese, and granted therefore that the machine doesn't understand either English or Chinese, still there must be something about me that makes it the case that I understand English and a corresponding something lacking in me that makes it the case that I fail to understand Chinese. Now why couldn't we give those somethings, whatever they are, to a machine?

I see no reason in principle why we couldn't give a machine the capacity to understand English or Chinese, since in an important sense our bodies with our brains are precisely such machines. But I do see very strong arguments for saying that we could not give such a thing to a machine where the operation of the machine is defined solely in terms of computational processes over formally defined elements; that is, where the operation of the machine is defined as an instantiation of a computer program. It is not because I am the instantiation of a computer program that I am able to understand English and have other forms of intentionality (I am, I suppose, the instantiation of any number of computer programs), but as far as we know it is because I am a certain sort of organism with a certain biological (i.e. chemical and physical) structure, and this structure, under certain conditions, is causally capable of producing perception, action, understanding, learning, and other intentional phenomena. And part of the point of the present argument is that only something that had those causal powers could have that intentionality. Perhaps other physical and chemical processes could produce exactly these effects; perhaps, for example, Martians also have intentionality but their brains are made of different stuff. That is an empirical question, rather like the question whether photosynthesis can be done by something with a chemistry different from that of chlorophyll.

But the main point of the present argument is that no purely formal model will ever be sufficient by itself for intentionality because the formal properties are not by themselves constitutive of intentionality, and they have by themselves no causal



powers except the power, when instantiated, to produce the next stage of the formalism when the machine is running. And any other causal properties that particular realizations of the formal model have, are irrelevant to the formal model because we can always put the same formal model in a different realization where those causal properties are obviously absent. Even if, by some miracle Chinese speakers exactly realize Schank's program, we can put the same program in English speakers, water pipes, or computers, none of which understand Chinese, the program notwithstanding.

What matters about brain operations is not the formal shadow cast by the sequence of synapses but rather the actual properties of the sequences. All the arguments for the strong version of artificial intelligence that I have seen insist on drawing an outline around the shadows cast by cognition and then claiming that the shadows are the real thing. By way of concluding I want to try to state some of the general philosophical points implicit in the argument. For clarity I will try to do it in a question and answer fashion, and I begin with that old chestnut of a question:

"Could a machine think?"

The answer is, obviously, yes. We are precisely such machines.

"Yes, but could an artifact, a man-made machine think?"

Assuming it is possible to produce artificially a machine with a nervous system, neurons with axons and dendrites, and all the rest of it, sufficiently like ours, again the answer to the question seems to be obviously, yes. If you can exactly duplicate the causes, you could duplicate the effects. And indeed it might be possible to produce consciousness, intentionality, and all the rest of it using some other sorts of chemical principles than those that human beings use. It is, as I said, an empirical question. "OK, but could a digital computer think?" If by "digital computer" we mean anything at all that has a level of description where it can correctly be described as the instantiation of a computer program, then again the answer is, of course,

yes, since we are the instantiations of any number of computer programs, and we can think.

"But could something think, understand, and so on solely in virtue of being a computer with the right sort of program? Could instantiating a program, the right program of course, by itself be a sufficient condition of understanding?"

This I think is the right question to ask, though it is usually confused with one or more of the earlier questions, and the answer to it is no.

"Why not?"

Because the formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless; they aren't even symbol manipulations, since the symbols don't symbolize anything. In the linguistic jargon, they have only a syntax but no semantics. Such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output.

The aim of the Chinese room example was to try to show this by showing that as soon as we put something into the system that really does have intentionality (a man), and we program him with the formal program, you can see that the formal program carries no additional intentionality. It adds nothing, for example, to a man's ability to understand Chinese.

Precisely that feature of AI that seemed so appealing – the distinction between the program and the realization – proves fatal to the claim that simulation could be duplication. The distinction between the program and its realization in the hardware seems to be parallel to the distinction between the level of mental operations and the level of brain operations. And if we could describe the level of mental operations as a formal program, then it seems we could describe what was essential about the mind without doing either introspective psychology or neurophysiology of the brain. But the equation, "mind is to brain as program is to hardware" breaks down at several points among them the following three:

First, the distinction between program and realization has the consequence that the same program could have all sorts of crazy realizations that had no form of intentionality. Weizenbaum (1976, Ch. 2), for example, shows in detail how to construct a computer using a roll of toilet paper and a pile of small stones. Similarly, the Chinese story understanding program can be programmed into a sequence of water pipes, a set of wind machines, or a monolingual English speaker, none of which thereby acquires an understanding of Chinese. Stones, toilet paper, wind, and water pipes are the wrong kind of stuff to have intentionality in the first place – only something that has the same causal powers as brains can have intentionality – and though the English speaker has the right kind of stuff for intentionality you can easily see that he doesn't get any extra intentionality by memorizing the program, since memorizing it won't teach him Chinese.

Second, the program is purely formal, but the intentional states are not in that way formal. They are defined in terms of their content, not their form. The belief that it is raining, for example, is not defined as a certain formal shape, but as a certain mental content with conditions of satisfaction, a direction of fit (see Searle 1979), and the like. Indeed the belief as such hasn't even got a formal shape in this syntactic sense, since one and the same belief can be given an indefinite number of different syntactic expressions in different linguistic systems.

Third, as I mentioned before, mental states and events are literally a product of the operation of the brain, but the program is not in that way a product of the computer.

-Well if programs are in no way constitutive of mental processes, why have so many people believed the converse? That at least needs some explanation."

I don't really know the answer to that one. The idea that computer simulations could be the real thing ought to have seemed suspicious in the first place because the computer isn't confined to simulating mental operations, by any means. No one supposes that computer simulations of a five-alarm fire will

burn the neighborhood down or that a computer simulation of a rainstorm will leave us all drenched. Why on earth would anyone suppose that a computer simulation of understanding actually understood anything? It is sometimes said that it would be frightfully hard to get computers to feel pain or fall in love, but love and pain are neither harder nor easier than cognition or anything else. For simulation, all you need is the right input and output and a program in the middle that transforms the former into the latter. That is all the computer has for anything it does. To confuse simulation with duplication is the same mistake, whether it is pain, love, cognition, fires, or rainstorms.

Still, there are several reasons why AI must have seemed and to many people perhaps still does seem – in some way to reproduce and thereby explain mental phenomena, and I believe we will not succeed in removing these illusions until we have fully exposed the reasons that give rise to them.

First, and perhaps most important, is a confusion about the notion of information processing: many people in cognitive science believe that the human brain, with its mind, does something called "information processing," and analogously the computer with its program does information processing; but fires and rainstorms, on the other hand, don't do information processing at all. Thus, though the computer can simulate the formal features of any process whatever, it stands in a special relation to the mind and brain because when the computer is properly programmed, ideally with the same program as the brain, the information processing is identical in the two cases, and this information processing is really the essence of the mental.

But the trouble with this argument is that it rests on an ambiguity in the notion of "information." In the sense in which people "process information" when they reflect, say, on problems in arithmetic or when they read and answer questions about stories, the programmed computer does not do "information processing." Rather, what it does is manipulate formal symbols. The fact that the programmer and the in-

interpreter of the computer output use the symbols to stand for objects in the world is totally beyond the scope of the computer. The computer, to repeat, has a syntax but no semantics. Thus, if you type into the computer '2 plus 2 equals?' it will type out '-4.' But it has no idea that '-4' means 4 or that it means anything at all. And the point is not that it lacks some second-order information about the interpretation of its first-order symbols, but rather that its first-order symbols don't have any interpretations as far as the computer is concerned. All the computer has is more symbols.

The introduction of the notion of "information processing" therefore produces a dilemma: either we construe the notion of "information processing" in such a way that it implies intentionality as part of the process or we don't. If the former, then the programmed computer does not do information processing, it only manipulates formal symbols. If the latter, then, though the computer does information processing, it is only doing so in the sense in which adding machines, typewriters, stomachs, thermostats, rainstorms, and hurricanes do information processing; namely, they have a level of description at which we can describe them as taking information in at one end, transforming it, and producing information as output. But in this case it is up to outside observers to interpret the input and output as information in the ordinary sense. And no similarity is established between the computer and the brain in terms of any similarity of information processing.

Second, in much of AI there is a residual behaviorism or operationalism. Since appropriately programmed computers can have input-output patterns similar to those of human beings, we are tempted to postulate mental states in the computer similar to human mental states. But once we see that it is both conceptually and empirically possible for a system to have human capacities in some realm without having any intentionality at all, we should be able to overcome this impulse. My desk adding machine has calculating capacities, but no intentionality, and in this paper I have tried to show that a system could

have input and output capabilities that duplicated those of a native Chinese speaker and still not understand Chinese, regardless of how it was programmed. The Turing test is typical of the tradition in being unashamedly behavioristic and operationalistic, and I believe that if AI workers totally repudiated behaviorism and operationalism much of the confusion between simulation and duplication would be eliminated.

Third, this residual operationalism is joined to a residual form of dualism; indeed strong AI only makes sense given the dualistic assumption that, where the mind is concerned, the brain doesn't matter. In strong AI (and in functionalism, as well) what matters are programs, and programs are independent of their realization in machines; indeed, as far as AI is concerned, the same program could be realized by an electronic machine, a Cartesian mental substance, or a Hegelian world spirit. The single most surprising discovery that I have made in discussing these issues is that many AI workers are quite shocked by my idea that actual human mental phenomena might be dependent on actual physical/chemical properties of actual human brains.

But if you think about it a minute you can see that I should not have been surprised; for unless you accept some form of dualism, the strong AI project hasn't got a chance. The project is to reproduce and explain the mental by designing programs, but unless the mind is not only conceptually but empirically independent of the brain you couldn't carry out the project, for the program is completely independent of any realization. Unless you believe that the mind is separable from the brain both conceptually and empirically – dualism in a strong form – you cannot hope to reproduce the mental by writing and running programs since programs must be independent of brains or any other particular forms of instantiation. If mental operations consist in computational operations on formal symbols, then it follows that they have no interesting connection with the brain; the only connection would be that the brain just happens to be one of the indefinitely many types of machines capable of

instantiating the program.

This form of dualism is not the traditional Cartesian variety that claims there are two sorts of substances, but it is Cartesian in the sense that it insists that what is specifically mental about the mind has no intrinsic connection with the actual properties of the brain. This underlying dualism is masked from us by the fact that AI literature contains frequent fulminations against "dualism"; what the authors seem to be unaware of is that their position presupposes a strong version of dualism.

"Could a machine think?" My own view is that only a machine could think, and indeed only very special kinds of machines, namely brains and machines that had the same causal powers as brains. And that is the main reason strong AI has had little to tell us about thinking, since it has nothing to tell us about machines. By its own definition, it is about programs, and programs are not machines. Whatever else intentionality is, it is a biological phenomenon, and it is as likely to be as causally dependent on the specific biochemistry of its origins as lactation, photosynthesis, or any other biological phenomena. No one would suppose that we could produce milk and sugar by running a computer simulation of the formal sequences in lactation and photosynthesis, but where the mind is concerned many people are willing to believe in such a miracle because of a deep and abiding dualism: the mind they suppose is a matter of formal processes and is independent of quite specific material causes in the way that milk and sugar are not.

In defense of this dualism the hope is often expressed that the brain is a digital computer (early computers, by the way, were often called "electronic brains"). But that is no help. Of course the brain is a digital computer. Since everything is a digital computer, brains are too. The point is that the brain's causal capacity to produce intentionality cannot consist in its instantiating a computer program, since for any program you like it is possible for something to instantiate that program and still not have any mental states. Whatever it is that the brain does to produce intentionality, it cannot consist in instantiating

a program since no program, by itself, is sufficient for intentionality.





# *Part 5: Can Machines Think?*

## **Artificial Intelligence and the Mind-Body Problem**

We encounter, in our modern lives, Artificial Intelligence (AI) more often than we would think. The spell-checker on our word-processors, Facebook's advertisements, driving directions on our phones/computers, some baby toys, speech-recognition software, and many other things all have AI built into them. They are made to make the machines intuitive to us and useful. These are examples of what is called 'weak AI' (a more precise definition coming later). When we think of AI, our minds are often plagued by thoughts of I, Robot, Star Trek, and other Sci-Fi stories. But, in the real world, can a machine, just lights and clockwork, have a mind? Can a computer be conscious? Can something like that really understand and really learn? Can there be a ghost in the machine?.

Those questions are all different versions of a seemingly simple question "can there ever be strong AI?" (again, more precise definition is later on). The philosopher John Searle in his paper *Minds, Brains, and Programs* was trying to answer just that.

Is it possible for a machine to think?

In our exploration of the Philosophy of Artificial Intelligence,<sup>53</sup> as it relates to the Mind-Body Problem, we will mostly be looking into the argument made by Searle in that paper and the replies to it, but I will be including examples to make it more relevant to today (as it was written in the 80s) and references to related more recent works.

## Weak Vs Strong Artificial Intelligence

Searle starts us off by going down the usual path for philosophy, making distinctions. We don't want to confuse simple forms of AI, like the kind found in a cell-phone, with complex forms, like the kind found in Data from Star Trek. Otherwise, if we were to think that all kinds of AI are the same, we would think that iPhones are iPeople. The kind of AI found in your phone and a lot of other computer-devices (including the one you are reading this on) is what we will call 'Weak AI'. This is:

A form of machine intelligence, focused on a small task or a narrow range of (interconnected) tasks. This is also called 'narrow AI'. The principle value or purpose of weak AI is to solve problems in a methodological and precise way which humans either don't have the brain-power or the time to do ourselves.

Weak AI simulates a person's thinking, typically in an ideal way, to get the best results. AI machines will not have the same unreliable aspects as a human's mind. For example, a suitably robust AI will not experience emotional fatigue which could result in a bias or missed factor, it will not (unless the programmer

---

<sup>53</sup>Selmer Bringsjord and Naveen Sundar Govindarajulu, "Artificial Intelligence," *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, fall 2022 (Metaphysics Research Lab, Stanford U, 2022).

built it in) experience the cognitive biases which we have seen in the past, and it will not experience becoming tired of doing the same task over and over again. For example, imagine that you are driving in an area which you are not familiar with and are using your GPS (in a phone or some other device) and you take a wrong turn. The GPS receives the data about your current position and from the algorithms figures out that you are not on the route. From this, it runs other algorithms to generate the fastest route to the destination given your current position. In a very short period of time, it generates the new route and the instructions for you to get to the destination. Imagine having your friend in the passenger's seat (riding shotgun) and having them be your GPS. The speed and accuracy of their directions would be nothing compared to that of the weak AI in the GPS.

Your computer or phone or what have you likely has many different compartmentalized weak AI programs in it, each activated for a particular task. The GPS AI is most likely not the same as the predictive-text AI. Human persons (a concept we will encounter in Module 9) aren't like that. Our intelligence is more holistic. The 'mind' which figures out your math homework is the same one which imagined the above example. This is where we get to Strong AI. This is:

A form of machine intelligence which is not focused on a small task or on a narrow range of tasks, but can handle just about any form of task which is thrown its way. Rather than merely simulating a person's thinking, the AI is, in fact, thinking. The principle value here is the same as that of any conscious human. Strong AI machines have minds.

Science Fiction is full of what we can use as examples for Strong AI. I have already used Data as an example, but also we have R2D2 from Star Wars, Sonny from i-Robot, Wall-E (from the movie with the same name), The Terminator (and many others if we include the plots of video games, like Cortana in

Halo). Searle has no problems with weak AI and, looking at the progress of technology, he was likely correct not to have any metaphysical qualms with it (ethical qualms are another story). But he has problems with the notion of strong AI. He does not think that strong AI is possible.

## Roger Schank's AI

Searle uses Roger Schank's AI as an example, because it's the one he's most familiar with. It should be noted that this AI is not special, but rather the basic way it works is found in AI even today, as I will explain in a moment. Schank's goal with this machine was to simulate how a person interprets a story when they're not given complete information. For example, take this story/question:

A man walks into a restaurant and orders a burger. It comes to him burnt to a crisp and he storms out angrily without paying or leaving a tip. Did the man eat the burger?

An answer to this question is not given in the story; it's not like you were asked "did he order a salad?" The answer is not spoon-fed to the machine or to us, rather we need to make logical conclusions given background information. In real life, many of the questions which we encounter do not contain all of the information which are necessary to answer them accurately. Most of the time, especially in the 'real world', we are going to need to make certain jumps in our reasoning based on background information. Such as our knowledge of normal human behavior.

More than likely, you answered the above question with something like "no, he didn't". The AI also generates the answer "no". This is because the AI in the machine was programmed or trained with cases involving reasonable human be-

havior and made predictions based on those assumptions. Next, take the following story (starting the same way):

A man walks into a restaurant and orders a burger. When it arrives, he is pleased with it and leaves a large tip before paying his bill. Did he eat the burger?

Yet again, the answer is not spoon-fed to the machine or to you. Rather, we need to use some background information. While it is possible that the man did eat the burger in the first case and did not in the second, this is far from likely. You probably answered the second question with "yes, he did". Similarly, the AI also says "yes". In this case, too, we are relying on our training and experience in dealing with real world situations, our past. If, for example, we had trained the machine or a baby with only experiences which were the opposite of the normal, then the baby and the machine would likely generate the opposite responses to us with our normal upbringing.

But, how do AI machines do this? Well, the bedrock level methodology has not changed much in the years, only getting faster and having larger data-bases for the relevant cases. The heart of it is a decision engine. This is some system of algorithms or if-then-else style statements which take in some input and produce some output. For example, we could have something like this going on:

If a person did not leave a tip, assume that they did not like the food. If a person did not like the food, assume that they didn't eat all of it.

The current rage in computer science involves using what are called neural networks<sup>54</sup> and machine learning (also called

---

<sup>54</sup>They are called neural networks because the connections between stimulus and response resemble the connections between neurons in the brain and they are strengthened in much the same way (by repeated exposure).

'learning algorithms'). In this case, the decision engine is generated by the machine itself. The AI starts off with a large database of cases along with an 'answer-key' of sorts. The easiest example would be a large number of pictures of hand-written numbers. The machine scans the picture and then makes a guess using some previously given (likely by the programmer) algorithm. Often, this is based on the contrasting pixels of the image and arrangements of similarly colored pixels. If it generates the correct number, great and it moves on to the next. If, on the other hand, the machine gets the wrong number, it adjusts the amount of 'weight' it gives some factor in the image (in this case the pixels) until it generates the correct answer. Doing this millions of times with millions upon millions of examples creates a decision engine which can accurately predict the answer for cases like the ones which it has been programmed to handle.

My case above using hand-written numbers is very oversimplified. For a more detailed account of what is really going on, the YouTuber 3Blue1Brown has several videos explaining this.<sup>55</sup> This methodology for machine learning and creating AI is not limited to cases like hand-written numbers and stories, rather it's found all over the place.

One shocking example of this in the real world is the case of Ashok Goel's Jill Watson.<sup>56</sup> Jill is an AI built to be a Teaching Assistant for Georgia Tech's MASSIVE online courses on programming artificial intelligence (meta, I know). Coming from experience, for massive courses like these, the Teaching Assistant is often bogged down by answering the same question hundreds of times a day. Though the questions are phrased differently (like how the numbers in the hand-written cases look different), they all boil down to the same answer. So, the professors at Georgia Tech collected a data-base of questions, sorted

---

<sup>55</sup>3Blue1Brown. But what is a neural network? Chapter 1, Deep learning. *YouTube*, Oct. 2017. [www.youtube.com/watch?v=aircAruvnKk](http://www.youtube.com/watch?v=aircAruvnKk).

<sup>56</sup>Talks, Tedx. A teaching assistant named Jill Watson; Ashok Goel. *YouTube*, Nov. 2016. [www.youtube.com/watch?v=WbCguICyfTA](http://www.youtube.com/watch?v=WbCguICyfTA).

them by type and gave Jill an answer key. After successfully generating correct responses for the initial data-base, Jill was given real-time questions being submitted in a real class (but not actually able to reply, she replied in a mirror-forum) and then graded by a real person on her responses. Once she had a 97% success rate (which is higher than some people I know), they let her loose into a real classroom forum. The remarkable thing is that very few people figured out that Jill was an AI. And even the ones who did, only did so because the class was on AI and they were already skeptical. In fact, the professor for the course needed to tell them that she was a machine. In the latest version of Jill which I know of, none were able to identify her as an AI.

## The Chinese Room Thought Experiment

Given Jill Watson, hand-writing recognition, and Roger Schank's AI, we have some basic knowledge of how AI machines make decisions and generate their answers, which is enough to get the point of this. The core commonality across all AI programs is a decision engine. How that engine is generated is not all that important (though using the machine learning and neural network systems seems to be the most efficient and most accurate), but what is important is that there's this engine. All Searle needs to make his argument work is that the machine takes input, runs it through an engine like Jill's, and then spits out an output.

Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese.

To help tie this together, in the case of Jill, the boxes of Chinese symbols is her data base of different answers. Similarly, in the case of Schank's AI, the possible responses to questions about stories are the Chinese symbols. The instruction manual is the core thing which we need to look into, this is the decision engine or sorting algorithm used by the computer to generate the answers. This can be written by a programmer with way too much time on their hands or through the kind of machine learning and neural networks I described before, that really doesn't matter. But what really matters is that all the machine is doing is, basically, crunching the numbers, running it through a bunch of 'if-then-else' style sorters to generate the answer.

If I were to lock you in the room with symbols/words not of your native tongue (one that you don't know), you would essentially be doing what the machine is doing, looking at the input, running through the instruction manual, and then giving the output. There would be no real understanding going on, no real learning in the process. If, on the other hand, I put you in a room full of symbols from your native tongue, there would be something more going on, something extra. There's actually interpretation happening. I don't know a lick of Chinese, but I



do know English, Latin, and the basics of a few other languages. In the case of Chinese, I would be just spitting out uninterpreted symbols. In the case of English or Latin, I would be interpreting the symbols, there would be intention or thoughts behind the answers given.

## Strong AI Claims and Replies

### Strong AI Claims and Replies

Generalizing off of the Chinese Room Thought Experiment, the proponent of Strong AI would make two claims:

1. The appropriately programmed AI (in this case the entirety of the Chinese Room) can be said, truthfully, to literally understand the applied case (in this case, the meaning of the Chinese symbols).
2. The machine and its programs explain the human ability to understand and reply the way we do in such cases (in this case, linguistic comprehension).

Searle thinks that neither the evidence nor the way in which even the most broad AI would function support these claims. Here is his reasoning:

In regards to the first claim, that the machine literally understands the applied case, it seems clear that the person in the Chinese Room is the same as the computer (or what have you) running the program. Though the person might generate answers indistinguishable from those generated by a native Chinese speaker, thereby be able to pass the Turing Test, there still would not be any real understanding, like we would have if an English speaker was put in the room with English symbols. This gives us the first argument:

If strong AI is possible, then the mere manipulation of symbols in a language would be enough to understand that language.

The mere manipulation of symbols in a language is not enough to understand that language.  
Therefore, strong AI is not possible.

For the second claim, that the machine explains the human ability to understand and reply in the way we do, Searle thinks that the programs described do not provide the sufficient conditions for understanding (if I am running the program, then I am understanding). This is because of core way in which they operate, formal symbol manipulation. It's possible to run the program without understanding (in the case of the Chinese Room). On the other hand, do the programs provide a necessary condition for understanding (if I am understanding, then I am running the program)?

The person who thinks strong AI is possible might go down that route, claiming that when I understand a story or dialogue in English, I am doing symbol manipulation, just of a far more complex and intricate kind. Searle is smart to point out that he did not show that this is false (claiming only that the machine does not give the sufficient condition for understanding). But, he does go a bit further, claiming that this is a truly incredible claim (an unbelievable claim). The plausibility of this claim rests on two claims. First, it is, in fact, possible to make a program indistinguishable from a native speaker of a language. Second, human persons are, at some level of description, programs. If you deny either of those, then you can't think that strong AI is possible.

In the case of the first claim, that it's possible for a machine to be indistinguishable from a native speaker, the jury is still out on that. All of the examples which I have encountered (and I will update this if ever I encounter a case otherwise) of a machine managing to trick a person into believing that it was a person, they were very sophisticated machines with a lot of exposure to various texts and responses but they were programmed to simulate a non-native speaker to play on our sympathies and make us hold them to a different standard than

we would a person who was clearly fluent in English and not prone to make simple errors. In order to prove that it is possible, it would take an advancement in computer science which we are currently awaiting, if it is possible at all.

In regards to the second claim, if one is a physicalist, then we could be willing to accept that human persons are on some level machines and functioning according to our programming (more on this in the Free Will Debate). However, there is a massive gap between the explanation in terms of the way our neurons are firing and the feeling, sense, of the world around us. Much like the Chinese Room, there is no room for the mental lives, understanding, feeling, in a purely physical explanation of the brain. The understanding must be over and above the matter in the brain (at least, that's what Searle is hinting at).

## Other Potential AI Formats and Rebuttals

Searle, during his time, presented this thought experiment around and got a few different replies to it, which he numbers and then gives the general regions where he got those replies. Many of the replies can be seen today in how some computer scientists are trying to make even more powerful AI. So, here we go (and these are fast spark-notes, for most there are far more detailed and extra rebuttals):

### The Systems Reply

While it's true that the person in the Chinese Room does not understand Chinese, the system as a whole does. Understanding should not be ascribed to the individual, but to the system as a whole.

For this, Searle's reply is quite simple, imagine that the person internalized all of the manual. The person has been locked in the room for so long that they have memorized the symbol

manipulation rules. Would that person understand Chinese? Just as before, there seems to be something missing, some intentionality, which marks the real understanding. We see this often in second-language classrooms, or at least the old-school ones which I dealt with for a time. Memorize the rules, some vocabulary (not knowing what the words mean), and then get trust into the language. Sure, the person might make the right replies in the right circumstances, but, essentially, the lights would be on and no one would be home.

## The Robot Reply

Rather than the program being in a immobile computer, suppose that we put it in a robot. The robot would not be 'bolted to the floor', but rather would be free to move around, eat, drink, make coffee, it would get sensory input from cameras and sensors, all of this would be controlled by the computer 'brain'.

Something akin to this has been made, and is one of the more successful examples of AI, not quite 100%, still has the problems. The difference between this machine and the first kind and Jill or Schank's AI is that they concede that the machine needs more than just formal symbol manipulation, more than just inputs and outputs, but needs to have the ability to interact with the world and develop a data-base from the real world experience. But, adding in the 'perceptual' and 'motor' qualities doesn't add anything to the base-level way that machines 'think'. For example, suppose that we replay the Chinese Room case, but this time the input is coming from a camera and sensors in a robot. The outputs would need to be more complex, obviously, but at the end of the day, it's still just a decision engine, there's no understanding in the machine.

## The Brain Simulator Reply

Suppose that we make a program which doesn't represent information which we have about the world, but rather simulates the actual sequence of neurons firing in a human brain. It takes in stories and simulates how the brain fires upon seeing the scripts and acts like the brain would command a body.

Now, where is the understanding in this system? Calling back to my A&P courses when I went to Community College, the brain is, simply put, an arrangement of neurons. The neurons fire and cause others to fire across the brain according to their arrangement. This is just a really, really, complicated decision engine, a really, super, complex set of if-then-else statements. It would only be simulating the structure of the brain, not the mind. If Non-Reductive Physicalism is true, then this, we could say is conscious... Maybe...

## The Combination Reply

What if we combined all three of them together. While each had a problem, namely versions of the Chinese Room, if we combined them we could get a way out. For example, what if we made a robot with a body indistinguishable from a human's. In the 'skull' cavity, there is a brain shaped computer. This computer would run a simulation of a human brain. We raise it as a human child, making no special circumstances for it (beyond those a parent would give their child), and so on. Imagine that the behavior of this robot is indistinguishable from a human person's. Surely, in this case we would say that it has intentional states (feelings and understanding).

In the real world, there have been attempts at this, with the

robot (though we can't simulate the brain yet) given a blank slate and raise as a child. The results of these cases were quite promising, but due to the limitations on technology, the brain power never really got far. Searle wants to point out a difference between appearance and reality. Certainly, if we come across a human on the street and they behave in the ways which we have come to expect people to behave, we would attribute to this human intentional states (feelings). But these are all appearances, NPCs in video-games have gotten quite good at appearing to have emotions and reactions, it's what makes contemporary video-games so awesome. For some interesting content on this, check out the The Uncanny Valley<sup>57</sup> (in this case, it's the reactions of the NPC, not the appearance, but the case still applies). But does an NPC really have those emotions? No, they don't, they are running a script. At the end of the day, the machine is still lights and clockwork. But a rebuttal to this reply leads to:

## The Other Minds Reply

How do you know that other people have intentional states (feelings)? I have first person access to my mental states, but other people's states are totally blocked to me. All I have to go on is their behavior. So, if you're going to attribute mental states to other humans (who can pass our intuitive tests), then by the same principle you must attribute them to a computer which can pass the tests.

This reply is, at first, very intuitive and seems to cut to the core of Searle's objections. Basically, the only reason that you think that I have a mind is because I look like you and I act a

---

<sup>57</sup>extrahistory. The uncanny valley - why more realistic characters look less human - extra credits. *YouTUBE*, May 2012. [www.youtube.com/watch?v=9K1Kd9mZL8g](http://www.youtube.com/watch?v=9K1Kd9mZL8g).

certain way. Isn't that enough for me to actually have a mind? However, as is often the case in philosophy, the beauty of a reply is only skin-deep. This reply is making a jump between an epistemic question (a question about what and how we know) and a metaphysical question (a question about what is actually the case). The core question for the strong AI proponent is not whether we know or believe that a machine has feelings, but rather whether the machine actually does have feelings. The study of knowledge is a separate question which we will cover in Module 6. In psychology and in our daily lives it is presupposed the reality and knowledge of other minds, just like how in the physical sciences and in our daily lives it's presupposed the reality and knowledge of the external world. These presumptions are not, necessarily, actually the case.

## The Many Mansions Reply

The core of Searle's argument makes an assumption, that AI is only about analog or digital computers. This just so happens to be the present state of technology, but what about the future? Whatever processes necessary for these intentional states, eventually, some system will be able to replicate it and that's what we will call strong AI. While we have weak AI today, nothing presupposes that some other means of making AI can't generate a strong one.

The only reply to this is that it moves the goalposts. The reason we are making AI in the way we are, aside from making life easier on us, is to hopefully explain some aspect of human intentionality. If we define strong AI as whatever gets us a really understanding and feeling thing, then making a baby is making strong AI. The core thesis, found in AI researchers even today, is that the mind is the brain, physical (remember that this is the core premise of physicalism!). If this thesis or

claim is reframed or redefined so that it's no longer physicalism, then the objections no longer apply and they no longer have a testable hypothesis.



## MODULE IV

*Do We Have Free  
Will? What is it?*

## *Part 6: The Free Will*

### *Debate*

The topic for this week is the notion of free will, what it is, whether people have it, etc. Stances regarding free will will influence your stances concerning problems in Neurology, Psychology, Criminal Justice, reward and punishment (praise and blame), morality (in general), physics, and many other areas of study. As a starting point, we will say that Free Will is whatever is necessary for moral responsibility. Other, more common notions of free will start with this as the beginning and then add things which they believe are necessary for such responsibility.

This means, basically, that if you aren't acting in accordance with your free will (or you don't have free will) then you aren't morally responsible for it. Others might think that you are and praise/blame you for the actions, but you aren't actually responsible for it, it just wasn't up to you. As a first example, suppose that some mad scientist puts a microchip in my brain and uses me as a remote control robot. She could make me do awful things and others, prior to knowing about my plight may hold me accountable for it, but in reality, I am not responsible for the actions made by my body.

That example brings us to an interesting point about moral responsibility/free will. It seems that there needs to be some

kind of control which the doer has over the action, something more than automated things. So, for example, take this case:

Sally is drinking a cup of coffee with a person she finds attractive (call it a date). They make a lewd comment and then, due to a strange muscle spasm, her hand flies forward as she's about to take a drink and splashes the coffee all over their white clothes. They will be furious of course and they will likely blame her for the strange event, but is she ACTUALLY responsible for it?

More often than not, regardless of whether you think that she should have splashed the coffee on them, you will think that she wasn't responsible in this case. So, what's missing? Now, let's look at this example, be sure to notice the difference:

Sally is drinking a cup of coffee with a person she finds attractive (call it a date). They make a lewd comment and then, in shock and anger, her hand flies forward as she's about to take a drink and splashes the coffee all over their white clothes. They will be furious of course and they will likely blame her for the strange event, but is she ACTUALLY responsible for it?

In this case, more often than not, you will say that she is responsible, for good or ill. So, what's the difference?

## Part 6.1: Free Will and (Moral) Responsibility

If there is no such thing as free will, there is no such thing as moral responsibility. This is because 'free will' is defined as whatever is necessary to have moral responsibility. So, we now need to ask various questions which can help us narrow the field concerning this responsibility:

| Question                                                                                                                | Likely Answer                                                         |
|-------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| Am I responsible for failing to do something I physically couldn't do?                                                  | If I can't do it, I can't be responsible for not doing it.            |
| Am I responsible for doing something which I physically couldn't not do (as in, it was impossible for me not to do it)? | If I had to do it (no choice), I am not responsible for doing it.     |
| Am I responsible for something outside of my control?                                                                   | If the thing is outside of my control, I can't be responsible for it. |

To help drive these points home, take these two examples, they are much like the examples which we have seen in this module already, so again, notice the difference:

You are you, as you are now, you can't fly, you can't run faster than a speeding bullet, you aren't bullet proof. A friend of yours is in New York and is outside of an orphanage on fire, with kids trapped inside. You are in Washington. Are you responsible for not saving the children?

As before, if I predict the ordinary intuition right, you will say that you aren't responsible for saving the kiddies. This is because it's just not within your power to save them. However, if we add in and change some things to the case, then we could think otherwise.

Last night, the chemistry labs on campus were busy, a meteor fell into a pool of some unstable compounds and you were exposed to the splash. You are not as you are now, you can fly, you can run/fly faster than a speeding bullet, you are bullet proof. A friend of yours is in New York and is outside of an orphanage on fire, with kids trapped inside. You are in Washington. Are you responsible for not saving the children?

If I predict your answer right, you will say “yes”. But, what is the difference, maybe Uncle Ben had it right, “With great power comes great responsibility.” Many people claim that the difference between the cases, as it regards moral responsibility is that in the first case, there’s nothing you can do, it’s not possible for you to help the kiddies but in the second, you can. This means that whatever is necessary for moral responsibility must have the requirement, built in, that you can do it. Some of you might have heard the phrase “ought implies can” which basically means that if something is your moral responsibility, if you are actually responsible for something, then you need to be able to do it. This also applies in reverse, if you can’t do something, then you aren’t responsible for it.

## Part 6.2: The Free Will Debate

Trying to answer questions like those is the basis for the free will debate. What makes those two cases different? Why do we say ‘no’ to the first example and ‘yes’ to the second example? And some of the possible ways of handling this were mentioned lightly above. Of the 8 possible stances which one can hold (taking the possible answers to various questions and using those to form stances), there are three which are actually palatable, able to hold their own. These stances are:

Hard Determinism

Libertarianism (not the political stance)  
Compatibilism

Both the Hard Determinist and the Compatibilist hold that Determinism is true. The Libertarian rejects Determinism. The debate concerns which of these stances is correct.

# *Part 7: Determinism, Incompatibilism, and Hard Determinism*

## **Part 7.1 Determinism**

This is the stance that every event is the result of preceding causes in accordance with the unchanging laws of nature. Basically, if some super amazing God-like computer knew all of the laws of nature and the state of the universe down to the smallest particles, it could predict with 100% accuracy the future. For example, take the muscle spasm coffee launch, the computer would have predicted it (prior to it happening). If the world was set-up differently (like in the second coffee case), the computer would have predicted the coffee splash there too. Also, there is nothing about the computer which is special, it is not determining your fate, it is just saying what will happen. The weather person saying that it will be sunny tomorrow does not make it the case that it will be sunny tomorrow. It is worth noting that your actions, choices, and thoughts are all events. This is pretty uncontroversial. So, if a stance says that deter-

minism is true, then it follows that all of the choices you make are just as determined by the past and the laws of nature as any other event, like an apple falling from a tree. This is a common area of confusion for students. Many think that Compatibilism includes provisions that everything else in the universe is determined except your choices. This is not true. Libertarianism is willing to make that claim. Compatibilism, as we will see, holds that determinism is true (your actions, thoughts, and choices are determined) but you are still responsible for your actions, under the right circumstances (which are more common than you would think). Actually, thinking about questions concerning our justice system, like 'why punish?' and 'how much?', from a deterministic mindset is why certain countries have a very low recidivism rate, and prison conditions are far better than in other countries. In those cases, they treat punishment as rehabilitation, removing the reason the person behaved as they did.

There is a lot of evidence in favor of determinism. In fact, the vast majority of the empirical sciences (biology, macro-level physics (not quantum mechanics (the jury is out there)), and chemistry to name a few) take determinism as a base level assumption. If two events have the same preceding causes (or relevantly similar preceding causes), then those events will be the same (or relevantly similar). We can see this in our daily lives as well. When we make a choice, there are many factors which enter into it, such as our past, our mental acuity in that moment, and our understanding of the case at hand. How we weigh or consider those factors have various causes as well, such as our past, dispositions (either nurtured into us or in our nature), and so on. All of those things, according to determinism, cause our choice with 100% certainty and, like the examples above, a super-computer, if it knew all of those factors, could predict, with certainty, the choices we make before we make them. The positive of this, then, is that if you want to alter your behavior, or the behavior of another, you need to think about why you or they made or will make those choices and



remove or change those factors.

## Incompatibilism

Incompatibilism is a family of stances regarding the relationship between free will and determinism. All stances in this family answer the question "Were determinism true, would we still have free will?" (or, in other words, "Were determinism true, would we still have moral responsibility?") the same way. If the stance says that there couldn't be responsibility in a universe where determinism is true, then the stance is Incompatibilist (free will and determinism are incompatible).<sup>58</sup> There are two, general, theories which fall into this family. These two are radically different in that they only agree about the answer to that question. These are Hard Determinism and Libertarianism.

## Part 7.2: Hard Determinism

Hard Determinism belongs to the incompatibilist family. This means that it claims that responsibility and determinism can't exist together. It also makes the positive claim that determinism is true, there is no randomness in the universe. It is called 'Hard' because it draws a hard line about the cases, takes the hard, difficult to swallow, parts of determinism to the extreme. As a result, from these two claims, it claims that there's no such thing as moral responsibility. We may hold each other responsible, but it is not a feature of the world (one might claim that morality is a useful fiction). There are several examples of this sort of thinking in the real world. For example, in a court case, suppose that the accused had a hard upbringing and up until that point lived a difficult life with few opportunities. They

---

<sup>58</sup>Not all determinists are incompatibilists. Also, for time-travel to be possible (as in traveling back in time), determinism must be true. I will not defend that claim here. As a side note, fatalism is determinism and if a model for the topology of time implies that time-travel is possible, then it implies fatalism; but it should be noted that there is some debate on this.

are in court accused of theft. The more you know about the accused and their upbringing, the less harshly you will want to hold them responsible or punish them because you quickly realize that in their mind, given their background and experiences, there wasn't another option.<sup>59</sup> Hard Determinism can use this fact to their advantage. We don't have the ability to have all of the factors of other people in our minds, but if we could, we would see that the responsibility we attribute to others is actually just a necessary result of the prior events, all stemming back to circumstances which the person had no control over. The control doesn't magically appear in some choices (because of determinism), so responsibility must just be a fiction our small minds concocted.

## An Argument for Hard Determinism

This is the basic argument for Hard Determinism, it relies on various assumptions which are quite commonly made in science, for example, that the laws of nature are constant (they don't change), that there aren't random events (even if there were, you would not get free will, in the sense of moral responsibility, but that's an argument for later), and, to a lesser degree, that the world is physicalist (not overly necessary, but makes it easier to argue). The bold text are the lines to the argument.

### 1. The past controls the present and the future.

This is the core tenant of determinism. It is basically that what happened in the past will tell you (assuming that the physical world is not random) what will happen in the present and the future. One way to think about it is how ordinary physical events, an apple falling, a rock rolling, etc. happen. Some change in the past determined a change in the present,

---

<sup>59</sup>This thought, taken in a slightly different direction, can be found in Martha Nussbaum's *Equity and Mercy*

and that determined a change in the future. Also, for a brain-bender (and people who speak languages which a different style of tense system than English will find this easier to grasp), what was the present is now the past and what was the future will be the present and thereby the past. This is the core tenant of determinism. It is basically that what happened in the past will tell you (assuming that the physical world is not random) what will happen in the present and the future. One way to think about it is how ordinary physical events, an apple falling, a rock rolling, etc. happen. Some change in the past determined a change in the present, and that determined a change in the future. Also, for a brain-bender (and people who speak languages which a different style of tense system than English will find this easier to grasp), what was the present is now the past and what was the future will be the present and thereby the past.

## **2. You can't control the past.**

This is likely the most intuitive of the lines of the argument. What happened happened, you can't change what happened, and make it different (and no, time travel stories where the characters change the past are not time travel stories or they are not possible, see the previous articles on this to see why). This might seem a little obvious; What has happened is set in stone, some might think that the future isn't written yet (though the determinist will disagree), but that notion relies on some aspect of the present being variable, random, or undetermined. The past can't be changed.

## **3. You can't control the way the past controls the present and the future.**

This is a special emphasis on one aspect of the previous line. Basically, it's here to remind us that we don't have control over how nature does its thing. The laws of nature don't change

because I want really hard and wish upon a star, the universe is horrifyingly indifferent. The core of it is that gravity will do its thing regardless of whether a human wills it one way or the other, and all other natural phenomena will work out the same regardless of us. With humans, we are just another cog in the machine. Remember that determinism extends to all events, which includes your thoughts, actions, and choices. You can't change who raised you, the past choices you made, the laws of nature and the past which lead up to the choice which you currently face. Those factors, according to determinism, determine the choice you will make in that moment. The 'up to you' aspects of responsibility doesn't appear.

#### **4. So, you can't control the present and the future.**

The word 'so' in this context is a marker for something derived from the other lines. And this line is derived from them. The only way out of this is to say that one of the previous lines is wrong. In this case, say that you can control the present/future, you will need to say that either you can change the past or change the laws of nature.

#### **5. If the way the present is and the future will be are outside of my control, you aren't not morally responsible for it.**

This is from a common reply to the question "am I responsible for things outside of my control?". When you think about it, being-in-control is required for responsibility. All I did was say that if the present and future are outside of my control, I am not morally responsible for it. The examples which I have given up to this point will drive home this point. For those examples of not being responsible for things outside of your control, think about this case:

A young person is driving to work, using the same back, forested, roads which they drive everyday. We could go into the causal factors which lead them to take these roads everyday, but that would be going in a different direction. The road bends and as they finish the corner, a deer leaps from the side of the road striking the front of their vehicle.

The various factors outside of this person's control, even if they were doing everything right up until that point, function as more than enough excuse to say that they weren't responsible for this. Relatedly, the nature of the laws of nature and the past also remove the responsibility which people could have for any action they take, any event which befalls them, because the factors leading up to them were outside of their control.

## **6. Therefore, you aren't morally responsible for anything (AKA, No Free Will)**

This is the conclusion, it is what all of the lines up until this point have been leading up to. If you accept all of the lines leading up to this point, you have to accept this one (really, no choice). People who accept this line are called Hard Determinists. The two main assumptions made are that Determinism is true and that responsibility requires some ability to do otherwise. But, there is another flavor of it called Libertarianism (not the political stance) which is incompatibilist, but thinks that determinism is false.

# *Part 8: What If We Denied Determinism?*

Basically, what would happen if we said that determinism was false? There are a few ways to deny determinism. Some posit that the laws of nature do have some randomness to them, pointing to some findings in Quantum Mechanics. They say that certain events on a quantum level ‘just happen’, there are no hidden variables. Two systems, they say, can be exactly the same, but behave differently. But how does this fair with responsibility? Well, it is not what is wanted. If there’s this randomness in the world, it makes our ideas of control even more elusive. Take for example, the muscle spasm coffee splash case. In that case, we say that she wasn’t responsible because it was outside of her control. Random events are always outside of our control. So, we don’t have free will for determined acts because we couldn’t do otherwise and we don’t have free-will for random acts because we don’t have control. But, there are some who deny this sort of reasoning, these are the Libertarians.

## **Part 8.1: Libertarianism**

Like the Hard Determinists, Libertarians are Incompatibilists. They hold, to say this again, that responsibility and determin-

ism can't play together. They disagree with the Hard Determinist, however, about whether determinism is true. They both hold that either determinism is true or there is moral responsibility (but not both). Hard Determinists say that determinism is true, so that removes responsibility (because it can't be both, this is not a formal fallacy, in this case). Libertarians say that determinism is false, so they get that there must be some indeterminacy and therefore responsibility. In particular, they deny the third line of the Hard Determinism Argument (you can't control the way the past controls the present and future). Although they are not determinists, Libertarians could be fine with physical events being deterministic, they want something extra beyond the laws of nature to intervene and give you control over the way the past controls the present and the future (there have been attempts to have a physicalist Libertarianism, but I argue against those attempts in some extra reading I wrote and provide). It posits some kind of substance dualism (typically, though there have been attempts otherwise, to have physicalism and libertarianism). The mental has control over the physical, within certain restraints from outside of nature: a "contra-causal" freedom, in which the mental is distinct from the causal order of nature, yet mysteriously able to alter it. In this way, we can deny the line in the hard determinist's arguments which says that past controls the present and future. According to these guys, we have some influence in there. The world is not deterministic, and we have moral responsibility. We could call that conception interventionist control.

## **An argument for Libertarianism**

This is an argument which I made up for libertarianism, sometimes when I make up arguments for a stance, for this class, I will intentionally give holes for you to find and exploit. This is not one of those cases. It starts from the basic assumption that there is moral responsibility, that morality is more than just a social fiction (it's an actual part of the world). We will

see arguments for and against this in Module 7. This particular argument is a shorter, simplified version of one I am giving in a paper.

- 1 There is moral responsibility (that is, a person is morally responsible for their actions, some or all).
- 2 If a person is morally responsible for their actions, then they must be able to do otherwise.
- 3 If determinism is true, then a person is not able to do otherwise.
- 4 So (from 1 and 2), a person is able to do otherwise.
- 5 So (from 3 and 4), determinism is false.
- 6 Therefore (from 1 and 5), there is moral responsibility and determinism is false.

The big area of contention is the second line, and that is worthy of a bit of a defense. Imagine the following case:

Suppose that a meteor crashed in a field near your home and an alien spore escaped filling the air, quickly infecting people's nervous systems. These spores grow and take over the mind of the host. Sometimes, the actions which the host would do are the same as those which the spores would cause them to do, however, when they are not, the spores switch up the nerves to make the host do as they would want. Since the host could not do otherwise, it would seem that they are not morally responsible for their actions (as even if they were their choice, the spores would stop them from making a different one).

To apply this more to the case at hand, the spores are the past and the laws of nature. We want to say that they are different than those, but the only reason for that is that the past and the laws of nature, according to determinism, have been with us forever, while the spores are a recent addition.



## *Part 9: Compatibilism and Soft Determinism*

The core similarity between the hard determinist and the libertarian is that they both think that determinism and free will can't play together, you get one or you get the other. That is, they are both incompatibilist. The opposite stance, compatibilism, says that they are compatible. One could hold that determinism is false but were it true, there would still be responsibility. This is all to say that one doesn't negate the other. Now, from my experience, versions of this stance are the most popular in philosophy at the moment. This distinction is useful because incompatibilism vs compatibilism is a choice or assumption we make prior to formulating or taking a stance in this debate.

### **Part 9.1: Soft Determinism**

Soft Determinism is a nicer, softer, interpretation of the implications of Determinism. This is the stance that determinism is true and that we have free will. Some people, causing some confusion, like to call this stance 'Compatibilism'. This stance is that the world is 100% determined but we still have some kind of control. I like to think of it as having your cake and

eating it too. Take this argument (shorter version of what was before):

The past controls the present and future.

You can't control the past.

Also, you can't control the way the past controls the present and future.

So, you can't control the present and future.

The Soft Determinist denies that we have no control over the future. Rather they say that we have control in virtue of being a cog in the machine. A common error which people have when they encounter this stance is that they think that the Libertarians, Hard Determinists, and the Compatibilists all mean the same thing when they use the term 'free will'. While it is true that all three mean 'whatever is necessary for moral responsibility', the Libertarians and the Hard Determinists hold that this responsibility requires that you physically be able to do otherwise. Determinism holds that it is physically impossible for you to do otherwise but it does not say that it was absolutely impossible for you to do otherwise. The Soft Determinists latch onto this sense. You are free because it was possible for you to do otherwise, even though you were physically guaranteed not to. For example, suppose that I am driving down the road and I come to a fork. I turn right. It was certainly be within my ability to turn left and drive that way just as much as it was within my ability to turn right. Determinism says that it was determined that I would turn right and I physically could not have taken a left, the fact still remains that I *could* have done it, making me responsible. I had control over the future, even though it was determined how I would use that control.

There are three ways that this intuition can be accounted for and each is stronger than the last.

## The Flash Definition

A subject acted freely if she could have done otherwise in the right sense. The subject could have done otherwise in this sense provided she would have done otherwise if she had chosen differently.

So, your action was free if you chose to do it; you would have done something different if you chose something different. Another explanation: For the libertarian, you get free will because you have control from outside of nature, (more than likely, through substance dualism, but that is a contemporary on going debate I am personally involved in). For the compatibilist of this stripe, you have control from within the laws of nature. Your freedom is in how you process the information and make your choices, though what choice you make is determined, you have free will when the choice is made in the right way. But, there is an issue for this, and I like to call it the mini-Martians problem, but the issue which I gave with the alien spores will work as well.

Imagine the invasion of the mini-Martians. These are incredibly small, organized, and mischievous beings: small enough to invade our brains and walk around in them. If they do so, they can set our modules pretty well at will. We become puppets in their hands. Of course, the mini-Martians might set us to do what we would have done anyhow. But they might throw the chemical switches so that we do quite terrible things. Then let us suppose that, fortunately, science invents a scan to detect whether the Martians have invaded us. Won't we be sympathetic to anyone who suffered this misfortune? Wouldn't we immediately recognize that he was not responsible for his wrongdoings? But, says the incompatibilist, why does it make a difference if it was mini-Martians, or causal agencies of a more natural kind?

Basically, there's no difference between the mini-Martians and the laws of nature, the more we learn about a person the more likely we are to think that they weren't responsible for their actions (which is why a Deterministic mind-set in cases of Crime and Punishment leads to such different results, you tend to see more rehabilitation rather than retribution). In the face of this, some philosophers have thought to rephrase the stance and add some more content, which gives us:

## Revised Definition

The subject acted freely if she could have done otherwise in the right sense. This means that she would have done otherwise if she had chosen differently and, under the impact of other thoughts or considerations, she would have chosen differently.

But this one, too, does have its issues. In this case, I can point to more trigger-warning-worthy real world examples, but

I will leave those aside. The core issue involves Bad Luck.

Although there is no randomness in this sort of world, we can still talk about ‘luck’ in the sense that something ‘just wasn’t in the cards’ or it was just the case that something would not happen. Sometimes, it is just not going to happen that the ‘right’ thoughts don’t arise.

Some philosophers like to associate freedom with understanding. We are free in so far as we understand. This is attractive to those who like political freedoms, like of speech and information. Including this in there will make the stance stronger. We need to figure out what “other thoughts and considerations” are. This is done by adding in that these are (1) accurate to the given situation and (2) available to the choicer.

## The Revised Revised Definition

This is the strongest of the bunch and best gets the contemporary views from this stance. It is a little more complicated, but when you apply it to cases in the real world where we don’t hold people responsible, it seems to fit well:

The subject acted freely if she could have done otherwise in the right sense. This means that she would have done otherwise if she had chosen differently and, under the impact of other true and available thoughts or considerations, she would have chosen differently. True and available thoughts and considerations are those that represent her situation accurately, and are ones that she could reasonably be expected to have taken into account.

# *Are Libertarianism and Physicalism Compatible?*

By Davis Smith

## **Abstract**

This paper concerns two seemingly unrelated topics, the Mind-Body Problem and the Free Will Debate. More particularly, I am worried about whether Libertarianism and Physicalism are compatible. First, I show that Libertarianism implies that there are actions such that the doer could have physically done otherwise and which the doer had control over. Next, I show that Physicalism implies that all actions are either deterministic or non-deterministic. And finally, I show that a Physicalist deterministic action eliminates the ability to do otherwise and that a Physicalist non-deterministic action eliminates the doer's control. These prove that Libertarianism and Physicalism are contradictory.

## Introduction

One would think that the more desperate two topics in philosophy are, the less impact a stance in one would have on the other. While this is true for some things, it does not seem to hold in the case of the Free Will Debate and the Mind-Body Problem. In this paper, I am going to argue that Libertarianism and Physicalism are not compatible.<sup>60</sup> This is to say that a universe with Libertarian free will cannot be a Physicalist universe. I do not claim that Determinism and Substance Dualism are incompatible<sup>61</sup> nor do I make the more robust claim that Libertarianism is self-contradictory,<sup>62</sup> rather, I am claiming only that Libertarian free will is contrary to Physicalism. This argument has 3 central premises, which the others prove/-support. First, if Libertarianism is true, then there are some actions such that the doer could have (physically) done otherwise and the action was within the doer's control. Second, if Physicalism is true, then for any action, that action is either deterministic or non-deterministic. And finally, if an action is deterministic, then the doer could not have (physically) done otherwise and if an action is non-deterministic, the action is not within the doer's control. These three premises, if they are accurate, lead to the necessary conclusion that if Libertarianism is true, then Physicalism is false.

---

<sup>60</sup>There have been a few papers in cognitive science and experimental philosophy which have shown that people who believe in Physicalism, more particularly, Reductive Physicalism, are less likely to believe in Libertarian free will. In other words, that there is a correlation between belief in Substance Dualism and Libertarian free will. See [Wisniewski](#) for one such study.

<sup>61</sup>I, personally, believe that they are compatible.

<sup>62</sup>This is a longer project which starts with a paper much like this one.

## Terms

For this phase of the argument, there are four (4) terms which I need to precisely define and clarify. For some, this will seem like simple review, but it is important that we are all on the same page. These terms are ‘Determinism’, ‘Physicalism’, ‘Free Will’, and ‘Libertarianism’.

### Determinism

An event A is deterministic if and only if, given the state of the world and the laws of nature prior to A, A will occur necessarily.<sup>63</sup>

Basically, for all possible worlds with the exact same laws of nature and they have the exact same past events leading up to the event, then the event will occur in all the possible worlds, in the same way and at the same time. The opposite of this is to say that an event is non-deterministic, and we can define this like so:

An event A is non-deterministic if and only if, given the state of the world and the laws of nature prior to A, A will not occur necessarily.<sup>64</sup>

This is not to say that the event will not occur, but rather it is saying that it is possible that the event will not occur. If an event is non-deterministic, then it is possible for the event

---

<sup>63</sup>Put in a more technical and precise way, this means that an event A (at t at w) is deterministic if and only if there is no possible world, w', such that the conditions of the world prior to t at w' and the laws of nature at t at w' are the same as those prior to and at t at w and A does not occur at t at w'.

<sup>64</sup>Put in a more technical way, we have that event A (at t at w) is non-deterministic if and only if there's possible world, w', such that the conditions of the world prior to t at w' and the laws of nature at t at w' are the same as those prior to and at t at w and A does not occur at t at w'.



not to occur in a world the same prior to it as a world where it does. There are two ways in which an event could be non-deterministic. First, the event could be the probabilistic result of a preceding cause or it could be completely uncaused. In the case of the former, the laws of nature would need to be in such a way that for at least some sets of circumstances, the probability of the event occurring is less than 1. The probability of the event occurring could be very close to 1, but if it is less than that, there is still a chance that the event does not occur. A totally uncaused event, on the other hand, would be one in which the preceding events in no way determine or make more likely that it occurs, such an event would be, for all intents and purposes, random.

With these two definitions, we can define ‘Determinism’ in the following way: Determinism is the stance that all events in the actual world are deterministic. If Determinism is accurate, then it follows that the laws of nature do not contain randomness and that the preceding events predict with 100% accuracy future events.<sup>65</sup> Some point to the seemingly random events at a quantum level as evidence against determinism, but some recent works<sup>66</sup> show that it is impossible to determine whether any notion of randomness can characterize the data in Quantum Mechanics. This means that appeals to such notions are going to, fundamentally, be based on intuitions and not the data.<sup>67</sup>

## Physicalism

For our purposes, though there may be other Physicalisms out there, this is a stance within the possible responses to the Mind-

---

<sup>65</sup>This way of defining ‘Determinism’ is used in several places, but most notably in [Popper](#).

<sup>66</sup>Jeffrey A. Barrett and Simon M. Huttegger, “[Quantum Randomness and Underdetermination](#),” *Philosophy of Science*, 2020, pp. 391–408.

<sup>67</sup>I will often use the phrase ‘deterministic universe’ which means that Determinism is true at the possible world in question. Similarly, ‘non-deterministic universe’ means that Determinism is not true at the possible world in question.

Body Problem. This stance is that all the objects/substances in the world are physical. There are no mental substances which do not reduce to the physical or which do not supervene on the physical. For our purposes, also, there is no reason to distinguish between Reductive and Non-Reductive Physicalism, as they both lead to the same relevant place.<sup>68</sup> Even in the case of Non-Reductive Physicalism, if a Physicalist universe is non-deterministic, then the indeterminacy is in the natural laws. Since mental events, at the very least, supervene on physical events, the choices we make (mental) are caused by changes in the physical nature of the brain. So, non-deterministic choices must be the result of a non-deterministic physical event.

## Free Will

For our purposes, I am defining ‘free will’ as whatever is necessary/sufficient for moral responsibility.<sup>69</sup> There are two ways which this can be interpreted:

1. If a person does an act freely, then they are morally responsible for that action.
2. If a person is morally responsible for an action, then they did that action freely.

The second interpretation seems to be more common. For example, many claim that if there’s no free will, then there’s no moral responsibility. The first notion, however, is the one which we will be using as this is seemingly what the Libertarians are after in the case of free will. Merely asserting that we have free will because Determinism is false does not guarantee that we

---

<sup>68</sup>I will often use the phrase ‘Physicalist universe’, which means that the possible world in question is one in which Physicalism is true.

<sup>69</sup>Others may fail to hold the doer morally responsible for the action or it may be the case that the degree of rectitude is inconsequential. Either way, the doer is still morally responsible.

have moral responsibility. This would be a fallacy. Libertarians, therefore, must hold the first conditional. They want free will because it grants moral responsibility.<sup>70</sup>

## Libertarianism

Libertarianism is an incompatibilist stance, meaning that it holds that Determinism and free will are not compatible. You could have one or the other, but not both. More particularly, it holds that Determinism is false, so there is free will and therefore moral responsibility. Since it implies that determinism is false, it states that there are some non-deterministic events. Some implied features of Libertarianism include:

1. To have moral responsibility, the person must have physically been able to do otherwise.
2. The events which cause the actions which we are morally responsible for, must be non-deterministic (because the doer must be able to do otherwise).

When it comes to the first feature of Libertarianism, this is the Principle of Alternative Possibilities<sup>71</sup>, but we need to be careful to include the term ‘physically’. Without it, the statement could be one which a Compatibilist would approve of. They would say that, sure, you could do otherwise, but you physically are not able to. Though the Frankfurt-style cases found in the literature are convincing to some that this is not an essential feature of moral responsibility,<sup>72</sup> I believe that they would be unconvincing to a sensible Libertarian.<sup>73</sup> The Libertarians would want the ability to physically do otherwise. With the

---

<sup>70</sup>Even those who wish to distance free will and moral responsibility will claim that free action is at least a sufficient condition for moral responsibility.

<sup>71</sup>See [Frankfurt](#).

<sup>72</sup>We will see a Frankfurt-style case in Part 1 of the argument.

<sup>73</sup>This is because the passive coercion in the Frankfurt-style cases is quite like the indirect causal pressures which the past and the laws of nature have

second feature, if the events which cause our actions are wholly deterministic, then it would not be possible for us to do otherwise, which means that, from the first feature, we would not be morally responsible for them.

## The Argument

This argument is broken into three parts, which come together at the end to show that Physicalism and Libertarianism are not compatible. The premises for the argument have four (4) conditional statements, each of which are proven in the relevant section. These are the same premises which I mentioned in the introduction. They are also the section titles.

### **Part 1: If Libertarianism is true, then there are some actions such that the doer could have done otherwise and the action was within the doer's control.**

Proving this line of the argument requires us to show two different things; both of which are derived from a basic understanding of 'moral responsibility'. From the very definition of Libertarianism, we have that we must have free will and therefore must be morally responsible for at least some of our actions. But what exactly does it mean to have moral responsibility for some of our actions?

It seems that there are at least two necessary features for a doer to be morally responsible for their action, and these are the two features of the above conditional. First, the doer must have been able to do otherwise. The doer must have had the ability to physically refrain from doing it or do something other

---

on our choices in a deterministic universe. If they reject the possibility of moral responsibility in a deterministic universe, then they would equally need to reject the possibility of moral responsibility in a Frankfort-style case. For an alternative argument, see [Widerker](#).

than what they did. For example, take this thought experiment to drive home the idea:<sup>74</sup>

Suppose that a meteor crashed in a field near your home and an alien spore escaped filling the air, quickly infecting people's nervous systems. These spores grow and take over the mind of the host. Sometimes, the actions which the host would do are the same as those which the spores would cause them to do, however, when they are not, the spores switch up the nerves to make the host do as they would want.

Since the host could not do otherwise, they are not morally responsible for their actions (as even if it were their choice, the spores would stop them from making a different one). When we think of cases where a person is forced or coerced to do something, then we are less inclined to hold them responsible for their actions. In the above case, the alien spores are passively forcing the hosts to do actions, so it seems clear that they are at least less responsible for the actions.

The other necessary feature for moral responsibility is a sense of control or up-to-us-ness. The doer must have chosen<sup>75</sup> to perform the action and they must have directed it to the outcome. For example, look at this case:

Suppose that you are at a dinner party having a discussion with various people and sipping wine. During a deep discussion on whether Libertarianism is compatible with Physicalism, you have a muscle spasm which launches the wine all over another's white cloths.

Whether the universe in this thought experiment is deterministic is not relevant. You have no control over what happens

<sup>74</sup>This is a case similar to the ones found in Frankfurt and Fischer.

<sup>75</sup>Making a choice, it would seem, needs to be both voluntary and with deliberation.

when you have a muscle spasm like this. That lack of control serves as a more than adequate excuse to relinquish moral responsibility for the destruction of another person's clothing.<sup>76</sup> One could also characterize this as an event which was not ultimately up to you. Your thoughts and attitudes, the reasons which you may have, did not enter the event. This means that, at least for the Libertarian, the ability-to-do-otherwise is not enough for moral responsibility. It is certainly the case that it's possible for you not to have that spasm, but the action must be up to you.<sup>77</sup>

**Part 2: If Physicalism is true, then for any action, that action is either deterministic or non-deterministic.**

This is straight forward, but I am the sort of person who does not like to leave anything without a proof. To start off, from the definition of 'Physicalism', we have that there are only physical substances. This is that there is only one kind of substance in the world and those substances are physical. This certainly removes the possibility of Substance Dualism, at least for the purposes of this argument.

The next line of the proof is that if there are only physical substances, then there are only physical events. Events require that there be objects or substances (one or more) which are involved in that event. An event A involves an object/substance B if and only if for all minimalist and accurate accounts the event A, the accounts contain reference to B. For a proof, try to imagine an event which does not have anything involved in

---

<sup>76</sup>For more on this, see [Austin](#).

<sup>77</sup>Here is a break-down of the argument: First, if Libertarianism is true, then people have free will. Second, if people have free will, then people are morally responsible for some of their actions. If people are morally responsible for some of their actions, then there are actions such that the person could have done otherwise and which the person was in control over. Therefore, if Libertarianism is true, then there are some actions such that the doer could have done otherwise and which was in the doer's control.

it. Even if there's quantum randomness where a particle just appears out of nothing, that particle is still involved in the event. Thus, if there are only physical substances, then there can only be physical events. Some might claim that there are mental events, but physicalism has it that those, at the very least, reduce to physical events or they are simply a different perspective on the physical event.<sup>78</sup>

Third, if there can only be physical events, then for any event, that event is either deterministic or non-deterministic. This can be derived using the law of excluded middle. Since there is no third alternative between being deterministic or non-deterministic, any event must be one or the other and the law of non-contradiction would have it that they cannot be both. From this, all we need is that all actions are events, which seems too obvious to deny.<sup>79</sup> As a tie between this part and the previous, Physicalist Libertarianism would say that there are at least some non-deterministic actions which we are, therefore, morally responsible for.

### **Part 3: If an action is deterministic, then the doer could not have done otherwise and if an action is non-deterministic, the action is not within the doer's control.**

This line of the argument has two halves. For both examples used to illustrate these conditionals, we can assume that they are in a Physicalist universe. For the first half of this part of the argument, this requires some quibbles about the exact meanings of the terms. The relevant sense of being able to do otherwise

---

<sup>78</sup>Also, a minimalist account of those mental events, if Physicalism is true, would only involve physical objects.

<sup>79</sup>As before, here is a breakdown of the argument: If Physicalism is true, then there are only physical substances. If there are only physical substances, then there are only physical events. If there are only physical events, then those events are either deterministic or non-deterministic. All actions are events. Therefore, if Physicalism is true, for any action, that action is either deterministic or non-deterministic.

is the one which the Libertarian uses. So, it follows naturally that a deterministic action lacks alternative possibilities.

The second half of this part of the argument is more involved. First, if an action is non-deterministic, then it is either a probabilistic effect of a preceding cause or it is totally uncaused. In the case of probabilistic events, though one event may be more likely than another, it is still, on a base level, random which of those events happens. For example, take the following thought experiment:<sup>80</sup>

A man is standing in front of a switch, there is a run-away trolley. If the man does nothing, then the trolley will kill 5 people, if they flip the switch, then the trolley will only kill one person. Since this is a non-deterministic universe, there is a 50% chance that they will flip the switch and a 50% chance that they will do nothing.

In this case, the core difference between it and other trolley problems is that there is this element of chance. Since this is a physicalist universe, the indeterminacy of the choice must be because of some manner of non-deterministic physical event. A minimalist account of the choice would not contain reference to the doer's reasons. The choice was not up to them.<sup>81</sup> The random nature of the choice takes the action outside of the man's control. In the case of totally uncaused events, the amount of control had by the doer becomes even more elusive. Take, for example, the following thought experiment:<sup>82</sup>

---

<sup>80</sup>The percent likelihood of the two different possibilities is, I would think, a worst-case scenario for the moral responsibility of the man in question. Other ratios of possibility are useable.

<sup>81</sup>It could be generalized that, since this is a Physicalist universe, the choice reduces to physical events and those must be probabilistic. On that level, the reasoning, the up-to-him-ness, is not present. A similar line of thought, though about artificial intelligence, can be found in [Searle](#).

<sup>82</sup>This case is similar in form to the ones seen in [Elzein](#), but the thought experiments there are used to show that the principle of alternative possibilities is important to moral responsibility and that the alternative possi-



One evening, Jones is sitting back to watch a little television when a quantum particle appears in his brain and then vanishes. This event causes a chain reaction which results in him choosing to begin growing edible mushrooms. Acting on this choice, he becomes very knowledgeable about the subject and eventually he discovers a new form of hypoallergenic penicillin, saving countless lives.

For ease of use, we can assume that the appearance of the particle is the only non-deterministic event relevant to the case.<sup>83</sup> Since the initial cause of the choice was completely random and the resulting events which lead to the discovery were the deterministic results of the initial event, we could hardly say that Jones was in control of the action and, thereby, morally responsible for saving the countless lives.<sup>84</sup> Along the same vein, going outside of the scope of this paper, if this were a Substance Dualist universe and the non-deterministic event occurred mentally, then Jones' choice still, ultimately, would not have been up to him.

---

bilities must be relevant to the case.

<sup>83</sup>Also, because other possible worlds are not relevant, we can assume that this is a physicalist universe.

<sup>84</sup>Here is a break-down of the argument: First, if an action is deterministic, then the person could not have done otherwise. If an action is non-deterministic, then it is either a probabilistic effect of a preceding cause or it's totally uncaused. If the action is probabilistic in this way, then the action was not within the person's control and if the action is totally uncaused, then the action was not within the person's control. Therefore, if an action is deterministic, then the person could not have done otherwise and if the action is non-deterministic, then the action was not within the person's control.

## Conclusion: If Libertarianism is true, then Physicalism is false.

So far, I have shown three (3) things. First, If Libertarianism is true, then there are some actions such that the doer could have done otherwise and that the doer was in control of that action. This is from the seemingly necessary features of moral responsibility. Next, I have shown that if Physicalism is true, then all our actions are either deterministic or non-deterministic. There is no third alternative. And finally, I have shown that if an action is deterministic, then it's not possible for the doer to have done otherwise and if the action is non-deterministic, then the action was not in the doer's control. From the second and the third, we have that if Physicalism is true, then for any action, the doer either could not have done otherwise or the doer was not in control. But the first part claims that if Libertarianism is true, then there are some actions where the doer has both of those features. This directly contradicts the result from Physicalism and therefore shows that if Libertarianism is true, then Physicalism is false. <sup>85</sup>

---

<sup>85</sup>Here the complete argument: If Physicalism is true, then there are only physical substances. If there are only physical substances, then there can only be physical events. If there can only be physical events, then for any event, that event is either deterministic or random. All actions are events. If an action is deterministic, then the person could not have done otherwise. If an action is random, then it is either a probabilistic effect of a preceding cause or it's totally uncaused. If the action is probabilistic in this way, then the action was not within the person's control and if the action is totally uncaused, then the action was not within the person's control. If Libertarianism is true, then a person has free will. If a person has free will, then they are morally responsible for some actions. If a person is morally responsible for an action, then a) that person could have done otherwise and b) that person is in control of the action. Therefore, if Libertarianism is true, then Physicalism is false.

# *Are Dualism and Libertarianism Compatible?*

By Davis Smith

## **Abstract**

This paper concerns two seemingly unrelated areas of Metaphysics, the Free Will Debate and the Mind-Body Problem. I will be arguing that Libertarianism and Substance Dualism are not compatible. The argument goes like so: First, if Libertarianism is true then there are some actions such that the actor could have physically done otherwise and which the actor had control over. Second, if Dualism is true, then there are three places where there could be the indeterminacy necessary for Libertarianism. Third, all three of these decrease the actor's control over their actions. Therefore, if Dualism is true, then Libertarianism is false.

## Introduction

Metaphysics is a very wide-reaching field and as such one would think that the conclusions one reaches in a far distant area would have little impact on the conclusions one could reasonably hold in a closer one. While this might be true for some areas of Metaphysics, it does not seem to hold for the Free Will Debate and the Mind-Body Problem. In this paper, I will be arguing that Libertarianism and Dualism are not compatible.<sup>86</sup> In a previous paper<sup>87</sup>, I showed that Libertarianism and Physicalism are contradictory, so here, using a similar methodology, I will show that Libertarianism and Dualism are contradictory. To give a brief overview of the argument: First, if Libertarianism is true then there are some actions such that the actor could have physically done otherwise and which the actor had control over. Second, if Dualism is true, then there are three places where there could be the indeterminacy necessary for Libertarianism. Third, all three of these decrease the actor's control over their actions. Therefore, if Libertarianism is true, then Dualism is false. We will start by being particular about the meanings of the terms central to this proof: Determinism, Dualism, and Libertarianism.

---

<sup>86</sup>I admit that this is counter-intuitive to some. There have been a few papers in Cognitive Science and Experimental Philosophy which point to a correlation between believing in Dualism and believing in Libertarianism. See [Wisniewski](#) for one such study. While I have no issue with Dualism, for the purposes of this paper, the issues I have are with Libertarianism.

<sup>87</sup>Davis Smith, "Are Libertarianism and Physicalism Compatible?", unpublished, presented at the 72nd Northwest Philosophy Conference. My main intent for this paper is to ultimately merge it with the Physicalism one, showing that Libertarianism is incompatible with the mind, whatever it may be.

## Terms

### Determinism

An event A is deterministic if and only if, given the state of the world and the laws of nature<sup>88</sup> prior to A, A will occur necessarily.<sup>89</sup>

Basically, for all possible worlds with the exact same laws of nature and they have the exact same past events leading up to the event, then the event will occur in all the possible worlds, in the same way and at the same time. The opposite of this is to say that an event is non-deterministic, and we can define this like so:

An event A is non-deterministic if and only if, given the state of the world and the laws of nature prior to A, it is possible for A not to occur.<sup>90</sup>

This is not to say that the event will not occur, but rather it is saying that it is possible that the event will not occur. There are two ways in which an event could be non-deterministic. First, the event could be the probabilistic result of a preceding cause or it could be completely uncaused. In the case of the former, the laws of nature would need to be in such a way that for at least some sets of circumstances, the probability of the event occurring is less than 1. The probability of the event occurring

---

<sup>88</sup> “Given the state of the world and the laws of nature” is meant to be neutral in regards to physicalism vs dualism.

<sup>89</sup> Put in a more technical and precise way, this means that an event A (at t at w) is deterministic if and only if there is no possible world, w', such that the conditions of the world prior to t at w' and the laws of nature at t at w' are the same as those prior to and at t at w and A does not occur at t at w'.

<sup>90</sup> Put in a more technical way, we have that event A (at t at w) is non-deterministic if and only if there's possible world, w', such that the conditions of the world prior to t at w' and the laws of nature at t at w' are the same as those prior to and at t at w and A does not occur at t at w'.

could be very close to 1, but if it is less than that, there is still a chance that the event does not occur. A totally uncaused event, on the other hand, would be one in which the preceding events in no way determine or make more likely that it occurs, such an event would be, for all intents and purposes, random.

With these two definitions, we can define ‘Determinism’ in the following way: Determinism is the stance that all events in the actual world are deterministic. If Determinism is accurate, then it follows that the laws of nature do not contain randomness and that the preceding events predict with 100% accuracy future events.<sup>91</sup> Some point to the seemingly random events at a quantum level as evidence against determinism, but some recent works<sup>92</sup> show that it is impossible to determine whether any notion of randomness can characterize the data in Quantum Mechanics.<sup>93</sup>

## Dualism

Dualism a stance in the Mind-Body Problem which says that rather than there being one kind of substance in the world, there are two, namely mental and physical. Under many normal interpretations of this stance, the physical substance is the physical body while the mental substance is the immaterial ‘soul’, this houses our feelings, sensations, thoughts, and the ‘what-it’s-like-ness’. Some Dualist theories hold that the mental and physical do not interact, there is no causal relationship between them. It assumes Determinism because it would be very contrary to our experience otherwise. We do not need to spend much time thinking about it for this project. Other Dualist theories, however, do posit that the substances interact. While

---

<sup>91</sup>This way of defining ‘Determinism’ is used in several places, but most notably in [Popper](#).

<sup>92</sup>See [Barrett and Huttegger](#)

<sup>93</sup>I will often use the phrase ‘deterministic universe’ which means that Determinism is true at the possible world in question. Similarly, ‘non-deterministic universe’ means that Determinism is not true at the possible world in question.

there is an open question about how this could be, we can assume that they do interact for our purposes.

## Libertarianism

Libertarianism holds that Determinism and responsibility for our actions are not compatible and holds that Determinism is false, so there is responsibility for our actions. From this, it states that there are some non-deterministic events. In general, being responsible for an action implies at least two aspects (according to Libertarianism):

1. The doer must have physically been able to do otherwise.
2. The events which cause the action which the doer is responsible for must be non-deterministic.<sup>94</sup>

When it comes to the first feature of Libertarianism, this is the Principle of Alternative Possibilities<sup>95</sup>, but we need to be careful to include the term ‘physically’. Without it, the statement could be one which a Compatibilist would approve of. They would say that, sure, you could do otherwise, but you physically are not able to. Though the Frankfurt-style cases found in the literature are convincing to some that this is not an essential feature of responsibility,<sup>96</sup> I believe that they would be unconvincing to a sensible Libertarian.<sup>97</sup> The Libertarians want the ability to physically do otherwise. Similarly, if the events which

---

<sup>94</sup>Mark Balaguer, who we will encounter later, claims that the indeterminacy must increase the control, a concept which we will explore in Part 1 of the argument. See [Balaguer](#)

<sup>95</sup>As presented in [Frankfurt](#)

<sup>96</sup>We will see a Frankfurt-style case in Part 1 of the argument.

<sup>97</sup>This is because the passive coercion in the Frankfurt-style cases is quite like the indirect causal pressures which the past and the laws of nature have on our choices in a deterministic universe. If they reject the possibility of moral responsibility in a deterministic universe, then they would equally need to reject the possibility of moral responsibility in a Frankfurt-style case. For an alternative argument see [Widerker](#).

cause our actions are deterministic, then it would not be possible for us to do otherwise, which means that we would not be responsible for them. Both of these, however, have reference to responsibility. Responsibility, as I will argue later, has at least two features which I believe contradicts these two features.

Most generally, Libertarians can be divided into two general teams. On one side we have the Agent-Causal Libertarians and on the other we have the Event-Causal Libertarians. Both teams agree that there are at least some actions which we are responsible for, they differ in how they think this functions. Agent-Causal Libertarians hold that the doer has the ability to be the first cause of a chain of events. These Libertarians hold that when we act freely, we are causing something to be without ourselves being caused.<sup>98</sup> I hold that Agent-Causal Libertarianism requires Substance Dualism, in other words, it is not possible to have this ability to be the 'first cause' without mental substances.<sup>99</sup> Event-Causal Libertarians take on a more metaphysically modest stance. These theorists generally have the best aspects of the compatibilists' theories but add the claim that there must be indeterminacy in the action. All Libertarians need to be able to show that the indeterminacy adds something to the control which is lacked in a Compatibilist model.<sup>100</sup>

---

<sup>98</sup>See Franklin 687 Some could see a similarity between this sort of idea and theological considerations. Some hold that God is the first cause of our universe and that God was not caused. Some also hold that God created man in His image. As a result, some could jump to the conclusion that God created us with the ability to be a first cause.

<sup>99</sup>The primary purpose of this paper is to show that all forms of Libertarianism, as I have constructed it, are incompatible with dualism. Agent-Causal Libertarianism seems a little more brazen about claiming that it requires dualism and the rejection of physicalism. If Agent-Causal Libertarianism does require dualism, then this paper serves as a proof that the stance is contradictory.

<sup>100</sup>I will argue in Part 2 that the indeterminacy detracts from control and thereby responsibility. The more indeterminacy we add to a given model, the less responsibility the agent has.



## The Argument

This argument is broken into three parts, which come together at the end to show that Dualism and Libertarianism are not compatible.

**Part 1: If Libertarianism is true, then there are some actions such that the actor could have physically done otherwise and had control over.**

Proving this line of the argument requires us to show two different things; both of which are derived from a basic understanding of ‘responsibility’. From the very definition of Libertarianism, there must be some responsibility for at least some of our actions. But what exactly does it mean to have moral responsibility for some of our actions?

It seems that there are at least two necessary features for a doer to be responsible for their action, and these are the two features of the above conditional. First, the doer must have been able to do otherwise.<sup>101</sup> As an example, take this thought experiment:<sup>102</sup>

Suppose that a meteor crashed in a field near your home and psychic alien spores escaped filling the air. Once some spores choose a host, they begin psychically implanting thoughts and take over their mind. Sometimes, the actions which the host would do are the same as those which the spores would cause them to do, however, when they are not, the spores implant thoughts which make the host do as they would want.

---

<sup>101</sup> And this must be a physical ability, not a metaphysical one, according to Libertarianism. So, the doer must have had the ability to physically refrain from doing it or do something other than what they did.

<sup>102</sup> This is a case similar to the ones found in Frankfurt and Fischer.

Since the host could not do otherwise, they are not responsible for their actions (as even if it were their choice, the spores would stop them from making a different one). When we think of cases where a person is forced or coerced to do something, then we are less inclined to hold them responsible for their actions. In the above case, the alien spores are passively forcing the hosts to do actions, so it seems clear that they are at least less responsible for the actions.

The other necessary feature for responsibility is a sense of control or ‘up-to-us-ness.’<sup>103</sup> For example, look at this case:

Suppose that you are at a dinner party having a discussion with various people and sipping wine. During a deep discussion on whether Libertarianism is compatible with Physicalism, you have a muscle spasm which launches the wine all over another’s white cloths.

Whether the universe in this thought experiment is deterministic is not relevant. You have no control over what happens when you have a muscle spasm like this and the lack of control serves as an excuse to relinquish responsibility for the destruction of another person’s clothing.<sup>104</sup> One could also characterize this as an event which was not ultimately up to you. This means that the ability-to-do-otherwise is not enough for responsibility. While it’s possible for you not to have that spasm, but the action must be up to you.

---

<sup>103</sup>This control requires that the doer had chosen to make the action and directed it to the outcome. Making a choice, it would seem, needs to be both voluntary and with deliberation.

<sup>104</sup>For more on this, see [Austin](#)

**Part 2: If Dualism is true, then there are three places where the indeterminacy could appear in an actor's action; either A) in the physical substance, B) in the mental substance, or C) in the interaction between the mental and the physical substances.**

In the previous part, we looked closely at what it takes to be responsible for our actions; the physical ability to do otherwise and control or 'up-to-us-ness'. For us to be able to do otherwise, in the Libertarian sense, the action or the events leading to the action<sup>105</sup> cannot be deterministic, meaning that there must be some indeterminacy in the actual process of making an action. It follows, then, that there are three possible places where the indeterminacy could occur: in the physical substance, in the mental, or in their interactions. Call these, in order, *physical-indeterminacy*, *mental-indeterminacy*, and *interaction-indeterminacy*. Of these, the indeterminacy therein could be totally uncaused or a probabilistic result of previous events. In the following sub-parts, I will show that each of these diminish or eliminate the doer's control.

**Part 2A: If there is physical-indeterminacy, then it decreases the control the actor has over their actions.**

In a previous paper, I argued that Physicalism and Libertarianism are incompatible on the grounds that physical-indeterminacy diminishes the doer's control or the 'up-to-us-

---

<sup>105</sup>There could have been an event in the distant past which was not deterministic which lead to the train of event leading to the doer's action, but such an event seems hardly relevant to the question of whether the doer could have done otherwise. For another example, suppose that one person had the ability to do otherwise while another did not. In their interactions, the free person caused the determined person's actions and the determined person could have, in a sense, done otherwise, but the determined person's actions were determined by the free person's.

ness'. This holds true here as well, for the same reasons. For example, take the following thought experiment:

A man is standing in front of a switch, there is a run-away trolley. If the man does nothing, then the trolley will kill 5 people, if they flip the switch, then the trolley will only kill one person. Since this is a non-deterministic universe, there is a 50% chance that they will flip the switch and a 50% chance that they will do nothing.

In this case, the core difference between it and other trolley problems is that there is this element of chance.<sup>106</sup> In a Dualistic universe, we can suppose that the mental substance had deliberated and, because of the past experiences and events in their life, choose to flip the switch. Since there is physical-indeterminacy, whether to flip the switch was not up to the man.<sup>107</sup> Generalizing this, if there is any physical-indeterminacy relevant to our actions, then our control is diminished by it.

In the case of a totally uncaused physical event leading to the action, take this thought experiment:<sup>108</sup>

---

<sup>106</sup>The percent likelihood of the two different possibilities is, I would think, a worst-case scenario for the responsibility of the man in question. Other ratios of possibility are useable.

<sup>107</sup>It could be generalized that in an extreme case like this, the doer's reasons and feelings do not play a part in the choice and thereby make it up to him. A similar line of thought, though about artificial intelligence, can be found in [Searle](#)

<sup>108</sup>This case is similar in form to the ones seen in [Elzein](#), but the thought experiments there are used to show that the principle of alternative possibilities is important to moral responsibility and that the alternative possibilities must be relevant to the case.

One evening, Jones is sitting back to watch a little television when a quantum particle appears in his brain and then vanishes. This event causes a chain reaction which results in him choosing to begin growing edible mushrooms. Acting on this choice, he becomes very knowledgeable about the subject and eventually he discovers a new form of hypoallergenic penicillin, saving countless lives.

For ease of use, we can assume that the appearance of the particle is the only non-deterministic event relevant to the case. Since the initial cause of the choice was completely random and the resulting events which lead to the discovery were the deterministic results of the initial event, we could hardly say that Jones was in control of the action and, thereby, responsible for saving the countless lives.

**Part 2B: If there is mental-indeterminacy, then it decreases the control the actor has over their actions.**

To illustrate this point, we can reuse the thought experiments from the previous part, with some minor alterations. The events in the mental substance could be either totally uncaused or probabilistic. For the probabilistic version, take the trolley problem case from before and think about it as indeterminism in the mental substance. In such a case, the person would not have control over their own thoughts and thereby their actions. They would be, in a sense, undisciplined. One could be faced with a choice where they are unsure what to do and deliberate about it. If the choice contains a hint of randomness, then that diminishes the control they had. Similarly, if the causal mental event is totally uncaused, then it would be very similar to the case involving hypoallergenic penicillin. The thought which spurs the action would be completely random and potentially radically out of character for Jones. This randomness further depreciates the responsibility Jones has.

**Part 2C: If there is interaction-indeterminacy, then it decreases the control the actor has over their actions.**

For the third and final place where this indeterminacy could be, take a look at the thought experiments once more. For probabilistic events in the interaction, it would be like a case where the mind chose to pull the lever and there is a 50-50 shot about whether the body would get the correct message. If the body, magically, got the message, then they would have gotten lucky in regards to doing the action. Often, when someone is trying to claim responsibility for an action or event, others will diminish that responsibility by claiming that they got lucky. If the indeterminacy required for the ability to do otherwise was in the interaction between the mind and the body, then it would be lucky that the body did what it was told. For totally uncaused events, this would be like the mental substance not giving an order to the body and the body magically getting an order. This too would be very troubling to how we could say that they are responsible for their actions.

**Part 3: So, if Dualism is true, then indeterminacy decreases the control an actor has over their actions.**

This is simple enough to prove. I have shown that, if Dualism is true, then there are three places where the indeterminacy required for responsibility, according to Libertarianism, could appear. Each of these actually diminish or eliminate the control the doer has over their actions. This means that if Dualism is true, then any form of indeterminacy (relevant to how our actions are taken) does not increase the responsibility one may have but rather decreases it.

## **Conclusion: If Dualism is true, then Libertarianism is false.**

To conclude this paper, I will outline what I have done. First, according to Libertarianism, Determinism is false and there are some actions which we are responsible for. These actions must be ones where we could have physically done otherwise (namely, there is a core aspect of indeterminacy) and those actions were within our control. Dualism allows for three places where the indeterminacy could appear: In the physical, mental, or in their interactions. I then went on to show that indeterminacy in the physical diminishes the control the doer has and it also does so when it is in the mental and also in their interaction. Libertarianism makes a central claim that this indeterminacy increases our responsibility for our actions, not decreases it. The implication from this is that if Dualism is true, then Libertarianism is false because in every place where the indeterminacy could be, it harms our responsibility, not help it.

## MODULE V

*Does God Exist? If so,  
why is there evil?*



## *Part 10: Arguments for the existence of God*

In philosophy, ‘God’ is given a standard three part definition, though some traditions include more features, others include less (I know of very few which include less). These features are:

Omniscient: All-knowing

Omnipotent: All-powerful

Omnibenevolent: All-good

So, these ultimately boil down to the idea that God is the ultimate being, there is nothing which She can’t do, nothing She doesn’t know, and She would never do anything morally wrong or unjust. The core question in philosophy of religion (I wish it had a different name, as this one can be misleading) is whether such an entity exists. A related question is whether we can prove it? Whether such an entity exists is a metaphysical question while whether we can prove it or know whether such a being exists before our deaths is a different kind of question, an epistemic one. There are many different proofs for the existence of God and all have at least one glaring problem.

For this section, we will be covering 4 different arguments for the existence of God and 1 argument against the existence of such a being. The arguments in favor of God’s existence are

The Design Argument, The Fine Tuning Argument, The First Cause Argument, and the Ontological Argument. I am willing to wager that if you believe in such a deity, the justification for the belief (likely) is some variation on the arguments seen in this class. For example, I love watching town-hall style school-board debates (when science textbooks are discussed) because I find arguments like the ones covered here 'in the wild'. Though, I will admit, it's very unlikely that I will ever see the last argument covered here in such an environment, which is a shame, as it's by far the strongest you are going to find, the rest are pretty bluntly seen.

It is very important that you know this is what philosophers mean by 'God', it may be different than the one you think of and many of the arguments for this being are from a monotheistic tradition. Talk of gods (plural) and arguments for them are not really found in the western tradition. But, I will also say, that all but the last argument can work for a pantheon of deities.

## The Design Argument

The design argument is actually a generic class of arguments which are also called **teleological arguments**. This style of reasoning goes as far back as Plato (I am serious), and in its most recent form, it can be found in the arguments for intelligent design (knowing this style of arguments and their flaws are helpful in the real world today because of this; regardless of the side you are on, you should know either your own flaws to defend yourself or the flaws in your opponent's stance to take them out). Plato argued that there was a designer of the universe because of the organization and structure of it. The odds of the universe coming about because of chance are very unlikely, so Plato and others believe that there must have been some designer. This is just an example to get the right thought-process in your head:

Suppose that you are walking through the forest, on a hiking trail, and you can across a computer, hooked to a running generator, mouse, keyboard, all on a desk. The computer is on and playing the classic Doom. Would you say that this appeared randomly, or would you say that someone built it?

More than likely, you will say that someone built it. However, the human body, with its cells and structures, is far more complicated than some computer. So, why do we say that it lacked a builder/designer, while the computer had one?

## The Argument

There are various different versions of this argument out there, and this is just my more formal, and improved version based on the spirit of it:

### **1. There are complex structures in the natural world.**

For evidence of this, just take a look outside. The trees, the mountains, the cells of the living animals running around, the living animals themselves all seem to be quite complicated, far more robust than a computer in a forest even.

### **2. If there are complex structures in the natural world, then they must have had a designer.**

This is the leap in the reasoning which you should look for in this particular argument, but there's another jump in the reasoning which is found across all of the arguments for the existence of God. This references back to the original intuition about the computer in the forest. If we are willing to say that there was a designer for this, then why not say that there was a designer for the even more complex living organisms and the world itself? There's (seemingly) no reason not to.

### **3. If there is such a designer, then that designer is God.**

This is the second jump in the reasoning and is the one which is found across the arguments for the existence of God. God is defined as the all-knowing, all-powerful, and all-good. If any being is capable of being the designer of the universe, then it most certainly would be God. However, as we will see later, this is not the only being capable of creating the universe...

### **4. If that designer is God, then God exists.**

**Therefore, God exists.**

## **Problems with the Design Argument**

### **Problem 1: The Weakness of the Analogy**

This is an attack on the first line of the argument. It only attacks why we think that it's true. The argument itself is valid. The first line seems intuitive, seems right, because of the examples used. The examples pointed to are things like watches and computers, very complex things which obviously needed to have some kind of designer. But the analogy between natural complexities and a watch seemingly has strength to it, but in actuality, the similarity is quite weak. There are many commonalities between two kinds of watches, but the commonality between the eye and a watch is that both are complex, there is nothing more to go on. When the analogy is weak, the conclusion from the analogy is equally weak.

Here are two arguments, so that you can see the weakness of the analogy. If you accept the first, based solely on the reasoning used, then you will also need to accept the second:

The brain is like a computer.

They both have complex, inter-working parts and it's very clear when one part fails to function properly.

Yet, it would be ridiculous to claim that a computer didn't have a designer.

So, claiming that the brain didn't have a designer is equally ridiculous.

Guns are like hammer.

They both have metal parts and could be used to kill someone.

Yet, it would be ridiculous to restrict the purchase of hammers.

So, restrictions on purchasing guns are equally ridiculous.

## Problem 2: Evolution

The Theory of Evolution gives us another reason to doubt the analogy given in the design argument. Although it does not disprove the existence of God, it just moves Him back a step, evolution gives us a way of saying that the eye and other natural complexities did not need a designer. The natural process of survival of the fittest (no, that does not mean survival of the best, but survival of the ones which best fit into their environment) shows how these complexities will come about without a designer.

## Problem 3: Limitation of the conclusion

This is shown by looking at the jump found in line 3 of the argument. This was:

The only thing which could have designed the complex structures in the natural world is God.

The point here is that even if we assume that there was a designer, we do not get that this designer was all-knowing, all-powerful, and all-good. More over, we do not get that there

was only one designer. For example, in the designing of a new watch, rarely is it one person, but a team of people working together. There is no reason to think that the divine designer of the human eye was the same who designed the brain, and there is no reason to think that either or both of those designers was any of the tripartite features ascribed to God. Following the same sort of structure as before, if you like the first argument, then you will also need to like the second:

Human bodies are remarkably complex and beautiful structures.

Such structures do not appear naturally without a designer/builder.

So, the Flying Spaghetti Monster, with His great noodly appendages, made them.

The pyramids are remarkably complex and beautiful structures.

Such structures do not appear naturally without a designer/builder.

So, someone, named Steve, all by himself, made them.

## The Fine-Tuning Argument

Even if we accept that there was no need for a designer in the minute machinations of human creation, there is another reason to think that there was some kind of divine player involved. At its core, the fine tuning argument points at the odds of some event happening. The odds of the world being as it is an suitable for human evolution and survival without the work of a divine architect are so astronomically small that it is more likely that it would have never happened at all. But, if you bring a divine architect into the mix, the odds are 100%. Since it did happen this way, the odds are far more likely that there was a divine architect. That Architect is God.

The Fine Tuning Argument could be seen as a variation or a more fine-tuned version of the Design Argument which we just covered. Although the Design Argument is quite ancient, the Fine-Tuning variation is relatively recent (as far as I am aware).

It mostly comes out of the advancements in the empirical sciences.

The argument goes like this, with the bold being the lines of the argument and the other being the explanation, note that when I extract an argument, I always make it as strong as possible, giving as much benefit of the doubt as possible. You don't want to kick a man when they're down:

### **1. The universe is suitable for human evolution.**

This should be pretty obvious. Given our knowledge of the world, we can say that we did evolve from some proto-human. Which means that the universe must be suitable for that event.

### **2. There are three ways that this could have happened: by chance, by necessity, or by design.**

This is supposed to cover all of the possible ways we have to account for the fact that the world has us on it. We could have been insanely lucky (1 out of  $10^{50}$  is far larger than the real chance by luck), it could be that the world (universe) just needed to be that way (could not have been otherwise), or something could have designed it for life.

### **3. The odds of this happening by chance are almost 0.**

This is basically saying that the odds are not in our favor. As I mentioned before, the odds are so indescribably small that the thought that it happened by chance should not enter our minds. (This is a particularly weak line, ripe for attack).

### **4. There is no reason to think that the laws of nature could not have been otherwise.**

In other words, there's no reason to think that the laws of nature are necessary. It seems that it's within the realm of possibility

that the laws of nature could be different. (Again, this is a weak line).

### **5. So, the world must be so suitable by design.**

This is the consequence of the previous lines, the second line gives us 3 options, the first 2 got taken out, so the third must be the right one. (This line is not open to attack, you need to take out a previous one).

### **6. If it happened by design, there must have been a designer.**

This is the same sort of move which we have seen before, relatively uncontroversial. If something was made, then there was a maker.

### **7. If there was a designer, then that designer was God.**

This is the God-jump which we have seen before. The reasoning here is essentially the same as the design argument.

### **8. So, that designer was God.**

### **9. Therefore, God exists.**

## **Problems with The Fine Tuning Argument**

With The Fine Tuning Argument, we can think of it as a more robust version of the Design Argument, so it does avoid some of the issues, for example it does not fall to Weakness of Analogy. In the text, rather than having the Theory of Evolution as an issue, we have The Lottery Objection and as with all the arguments for the existence of God, we will have Limitations



on the Conclusion. There are three other objections which were not covered in the textbook, which I have added.

### **The Lottery Objection**

The core different objection to this is concerning its very reasoning. Very unlikely events can still happen. Suppose that you are one of millions who bought a lottery ticket to win several million dollars. Now, one of the tickets will win, the odds of it being yours are small, but that is true for all of the other tickets. Suppose that you pick a planet in the universe at random (your lotto numbers), the odds of that planet being one which can support life are small, but that is true for any planet you could have picked. Given the sheer number of planets, one or more will support life. Whatever creatures evolved on those planets may think that their position was put there by a designer too.

### **Limitation of the Conclusion**

As with the Design Argument, there is a jump in the reasoning which is unsupported. In this case, the jump is from there being something which tuned the world/universe to be suitable for evolution/intelligent life to that thing being God. To illustrate, just as before, take the following arguments, if you accept the reasoning in the first, then you must also, by the same logic, accept the second, but they are not compatible:

The world is suitable for human evolution.

The odds of this happening without a divine architect are next to 0.

The odds of this happening with a divine architect are 100%.

So, there is a divine architect. That divine architect is God.

Therefore, God exists.

The world is suitable for human evolution.

The odds of this happening without a divine architect are next to 0.

The odds of this happening with a divine architect are 100%.

So, there is a divine architect. That divine architect is a teenager doing a science fair project.

Therefore, a teenager doing a science fair project exists.

## Scientific Progress

The basis for the Fine Tuning Argument is that there is currently no explanation for the constants of nature being so precise. However, this is an issue only currently. In the history of humanity in general, there have been many phenomena for which we lacked explanations and which were eventually explained away (such as the weather, the sun, stars, and natural disasters). The empirical sciences may, one day, find an explanation for these constants which is completely satisfactory.

## Exotic Life

Another area of contention for the Fine Tuning Argument is that it says that it's suitable for human life. While humans are quite remarkable beings, why should we think that the universe was specially made for us? It's at least conceivable that in a universe where the constants are not suitable for human life, there could be exotic life which is remarkably different than our own.

There are limits to this response however, such a universe would need to be set up to overcome the second law of thermodynamics, which is easier than some would think (there's more possibilities for that than normally presented).

## The First Cause Argument

First Cause Arguments are also called **cosmological arguments** as they, rather than looking at how the world is, they look at the cause and effect relations which led to the world as it is. Everything in the natural world has a cause. All of these causes in turn have a cause. There are things in the natural world. So, there must have been a first cause, something not in the natural world, which created this chain (the cause of the big bang is the new standard place to put it). This first cause was God.

### 1. Everything has a cause other than itself.

This seems sort of self-evident, but might be worth explanation. Think of where you came from. Personally, my existence was caused by a drop in copper prices coupled with the series of events which lead my parents to meet at a desert party while my dad was a student at ASU and them moving to California. I can go farther, my father (whose story I am more familiar with) was caused by my grandfather and grandmother meeting, along with the events leading up to that, and so on. No where in the chain of cause and effect (unless backwards time-travel is possible/actual) will you find a case where something, directly or through transitivity, causes itself. More over, all things are a part of a causal chain like this.

## 2. Everything which is a cause has itself a cause other than itself (directly or indirectly).

This is mostly an expansion/explanation of the first line, meaning that there are no existing things which weren't caused to exist by something previous to them in a chain. As I noted in parentheses above, there is an issue if you allow for backwards time-travel. In such a case, it would be possible for a thing/person to cause themselves to exist. For example, take this case from the Futurama episode "Roswell that Ends Well":

Philip Fry travels back in time to Roswell, New Mexico in the year 1947, where his 'grandfather' was stationed. As events play out, Fry ends up causing the death of his 'grandfather'. But, Fry does not fade away to nothing, and though some faulty reasoning, comes to believe that the woman he is consoling is not in fact his grandmother (she is, though), and Fry ends up sleeping with her. As a result, she becomes pregnant and Fry becomes his own grandfather.

Here, we have a case where, thanks to time-travel, a person was their own cause. This can be an issue for this argument, but it's better as the 'Not a Proof' objection in the next page.

## 3. There are things which exist.

This is another one of those 'no duh' sort of lines. Basically, it points out that these non-looped causal chains are in the world and need some kind of explanation. So, using an interesting jump in the reasoning (which will be pointed out later), we get the next line:

#### **4. So, there must have been a first cause to kick start the chain.**

As always, the word 'so' indicates that we have an intermediate conclusion, this means that it's supposed to follow from the previous lines, whether or not it does is a different question. Basically, the reasoning here is that since there are things in the world, and everything needs to have a cause other than itself, there needs to have been something to start it, something which got the ball rolling.

#### **5. If there was a first cause, that cause was God.**

This sort of jump should look familiar, as it is found in both the Design and the Fine-Tuning Arguments. This jump is often found when people look at the big bang and ask "OK, what caused that?" When they find they get something like "hey, we don't know" they jump to God. This jump is just as exploitable as it was in the previous arguments. And this is where, from this line and the previous, we get the conclusion:

#### **6. Therefore, God exists.**

### **Problems with the First Cause Argument**

#### **Problem 1: Self-contradictory**

In some versions of the argument, they are not careful with their phrasing and claim that *everything* is said to have a cause other than itself, so God must have had a cause and there could not be a first cause. For example, take this argument:

1. If it exists, it has a cause other than itself.
2. If something is a cause, it has a further cause.
3. God exists.
4. So, God must have had a cause other than Himself.

5. If there was such a cause, that cause was Meta-God.
6. If that cause was Meta-God, then Meta-God exists.
7. Therefore, Meta-God exists

All I did here was apply the exact same reasoning which the First Cause argument uses to get that God exists, but applied it to God Himself (with the same restrictions). Since I am not a fan of taking on the letter of an argument and I prefer to attack the spirit of it, I have made this argument stronger by limiting the 'things' to objects in the natural world. Many of the versions found in popular culture have this error and this saddens me because it's such an easy fix:

1. Everything in the natural world has a cause other than itself.
2. Everything which is a cause in the natural world has itself a cause other than itself.
3. There are things in the natural world.
4. So, there must have been a first cause from outside of the natural world to kick start the chain.
5. If there was a first cause from outside the natural world, that cause was God.
6. Therefore, God exists.

This version, also, has its own problems which are in the very spirit of the argument, some of which we saw in the explanation of the argument.

## **Problem 2: Not a proof**

This argument makes two background assumptions each of which can be used to show that it does not take all relevant

possible cases into account, and thereby not really work as a proof. First, the argument assumes that the series of cause and effect can't go back infinitely. But, this assumption requires its own argument to prove and may be intuitive because of our limited understanding of the earliest moments in the universe. This under-supported assumption does imply that there must have been a first cause. But, if you look at that assumption, there's no reason to think that.

More over, the second line of the argument claims that nothing (at least in the natural world) can, directly or indirectly, cause itself to exist. But, in the previous page, I gave a seemingly possible case, assuming that you can have time-travel, where circular causation occurs. The chain doesn't, in a sense, continue back indefinitely and it does not have a first cause. Applying this to the universe itself and without the need for time-travel, it could be the case that all of time is circular, where the cause of the 'first' moment of the universe was the 'last' moment of the universe. This can be confusing, as in a circular model of time, ordering moments in terms of 'first' and 'last' is misleading, there are no such moments.

### **Problem 3: Limitations of the conclusion**

Just like with the other arguments for the existence of God, even if we accept all of their reasoning, there is still the jump. There is no reason to think that the first cause was God or any being with one or more of the tripartite features of God. It could have been a really evil kid playing with the equivalent of an ant farm. The universe could have been his ant farm. To illustrate this using other examples, take these. As before, if you accept the reasoning in the First Cause Argument, then you must also accept these, which are equally strong arguments:

If it exists in the natural world, it has a cause other than itself. If something is a cause in the natural world, it has a further cause.

Some things exist in the natural world.

So, there must have been a cause from outside of the natural world to start it.

If there was such a cause, that cause was random fluctuations of quantum strings.

If that cause was random fluctuations of quantum, then random fluctuations of quantum strings exists.

Therefore, random fluctuations of quantum strings exists.

If it exists in the natural world, it has a cause other than itself. If something is a cause in the natural world, it has a further cause.

Some things exist in the natural world.

So, there must have been a cause from outside of the natural world to start it.

If there was such a cause, that cause was The Flying Spaghetti Monster.

If The Flying Spaghetti Monster was the cause, then it exists.

Therefore, The Flying Spaghetti Monster exists.

The left most argument, if I remember correctly, is roughly one given by Steven Hawking when there was a debate about this topic between physicists and theologians, but this was many years ago.

## The Ontological Argument

To start us off, I will define a special term, which you might have encountered before. **Ontology**, properly speaking, is an area of philosophy dealing with existence. It is the study of being, becoming, and some features of reality relating to existence. I spent the majority of my undergraduate work dealing with these sorts of questions as it relates to time and also to composition. Sometimes, Ontology is called "Metametaphysics", but Metametaphysics tends to be more of Meta-ontology.



This term is also used in Computer Science, but there it's the formal representation of categories within a data structure. They use this term, I think, because much of the work in Ontology in Philosophy does relate to the representation and naming of objects in the world.

As the name implies, **ontological arguments** are arguments which deal with the existence of something by looking at the very nature of that thing. These sorts of arguments are very different from those we have seen, and they are typically not employed in pop-culture. They do not use evidence about how the world is, was, or will be, to show that the thing must exist. This was the major flaw in the previous arguments, because we can just counter those assertions with other possible explanations. Not relying on evidence, in this context, makes the arguments significantly stronger. Also, these kinds of arguments are a bit harder to wrap your head around, which is probably why they aren't being used in school-board debates.

## Anselm's Ontological Argument

The oldest, and one of the strongest, ontological arguments was written by St. Anselm. Anselm was born in 1033 CE and died in 1109 CE. He was proclaimed a saint in 1163 CE and his feast day is April 21st. As far as I could find, he is not the patron saint of anything.

Anselm starts by defining God in a different way than the one which we have seen. Thus far, we have had God be whatever has these three features: All-Knowing, All-Powerful, and All-Good. But, Anselm defines God as:

"That being than which nothing greater can be conceived."

This definition does rely on a few things, which weaken the letter, but not the spirit of the argument (such as the powers of human imagination), so we will make it a little stronger by saying that God is:

”The being such that there is nothing can be greater than it.”

Essentially, we are saying that God is the greatest possible being, in all respects. But, first off, are we talking about the same thing? Some could easily argue yes. When it comes to power, nothing can be more powerful than God (as Anselm defined the term), so God is all-powerful. Nothing can know more than God, so God is all-knowing. With goodness, we run into an interesting point though. Anselm thinks, as many do, that evil is the absence of good, so ‘the greatest evil’ is a misnomer and it should be ‘the least good’. This means that God can’t be the greatest evil possible, but is the greatest good.

Remember that Ontological Arguments look at the very nature of a thing to show that it must exist. Does being the greatest in everything imply that God exists? According to those who like this argument, YES! And here is their argument:

### **1. God is the greatest possible being.**

This is the definition we are working from, remember that this gets us the being which we have discussed before, but also gets us some other features which may have been assumed by some of you before, such as omnipresence.

### **2. Any existing being will be greater in some respects than a non-existing being.**

This one is a little different, so I will use certain examples to help. I exist and because of this, I have certain powers and abilities which non-existent things lack. For example, I can interact directly with physical objects, while non-existing things, like fictional characters can’t do that, their powers are limited to the will of the author and the minds of the reader. So, in at least some respects, I am greater than a non-existing thing, in virtue of my existence. This does not just apply to me, but also to any existing thing, I would just need to fiddle with the examples.

### **3. The greatest possible being cannot have anything greater than it in any respects.**

As we have defined it, the greatest possible being is the greatest in all respects. As such, no being, existing or not, can be greater than it in any respects. But, as a tie in, remember that any existing thing will be greater than a non-existing one in some respect.

### **4. So, if the greatest possible being does not exist, then there are things which are greater than it in some respects.**

If existing things are always greater than non-existing ones in some respects, then if the greatest possible being doesn't exist, there would be something, namely an existing thing, which has powers, abilities, positive features which it lacks. But, the greatest possible being having something greater than it should seem contradictory.

### **5. So, the greatest possible being exists.**

This comes from the fourth line. The seeming contradiction is actual (or defensible). Assuming that the greatest possible being doesn't exist gives us a contradiction, so long as there are existing things (which should be obvious). If something leads to a contradiction, then we know that it can't be true. So, the greatest possible being must exist.

### **6. Therefore, God exists.**

Remember how we defined God, God is the greatest possible being. What is true for one is true for the other, they are the same thing.

## Problems for the Ontological Argument

### Problem 1: The Perfect Island

This was actually the inspiration behind my versions of the arguments seen in the Limitation of the Conclusion sections. It is easy to imagine the absolutely perfect island. Perfect weather, wildlife, beach, and so on (think Hawai'i without the tourists). But, that island would be more perfect if it existed (same reasoning as before). This island is the most perfect island, therefore, the perfect island must exist. We can do that for the perfect taco, the perfect wife/husband, or anything, the perfect version of it must exist. Now I have to ask, where is my perfect island?

This is the Ontological Argument's equivalent of Limitation of Conclusion, but you need to be a little more careful about the counter examples used. Sometimes, the example won't work because of the nature of the thing, for example, the perfect fictional character can't be used for this because the perfect fictional character can't exist, because then it would not be fictional. Take for example, these two arguments, which follow the same reasoning as the Ontological Argument:

Alex is the greatest possible significant other.

Any existent being is greater in some respects than non-existent beings.

The greatest possible significant other can not have anything greater than it in any relevant respects.

So, if the greatest possible significant other does not exist, then there are things which are greater than it in some relevant respects.

So, the greatest possible significant other exists.

Therefore, my significant other, Alex, who lives in Canada, exists.

God is the greatest possible being.

Any existent being is greater in some respects than non-existent beings.

The greatest possible being can not have anything greater than it in any respects.

So, if the greatest possible being does not exist, then there are things which are greater than it in some respects.

So, the greatest possible being exists.

Therefore, God exists.

## Problem 2: Existence is not a property

Take for example, 'unicorn'. The definition of that word is 'a horse with one horn on its head'. Now, if I said 'unicorns exist', I am not adding another property to unicorns. For anything to be a unicorn, it must first exist. But the concept of unicorns does not change whether or not there are unicorns.

Applying this to the argument, we notice that it treats existence as just another property. Something which a thing is greater for having. But this is a mistake, existence is often seen as a precondition for having properties. We can have the concept of the greatest possible being, but that does not entail that such a being exists.

In grad-school, I did several presentations arguing against the idea that existence is not a property and also against that existence is a precondition for having properties. But that was

really advanced stuff and I will not go into depth about it here.

### **Problem 3: Evil**

This is our segue into the next topic of this module, which is the problem of evil. If we look at the world, there is a HUGE amount of pain and suffering in it (AKA evils). If God does actually exist, then there would not be these evils (because He is all good). So, God must not exist.

# *Part 11: An Argument Against the Existence of God*

The **Problem of Evil**, which I abbreviate as POE, is the family of arguments against the existence of God which we will be discussing. This is one of those arguments where what doesn't kill it makes it stronger. As a result, there are many different versions of it, each with their own special objection (explanations for how God can exist despite the argument). The objections to the arguments for the existence of God were not arguments against the existence of God, rather they are ways that the argument for the existence of God is flawed. You do not want flawed arguments. For this class, we will deal with, in general, two kinds of POE, which are separated by the kind of evil they use.

First, we have what I'll call **man-made evil**. These are also sometimes called 'moral evils'. Man-made evils are the sufferings which people cause, the 'bad' things we do. Do not get hung up on the use of the word 'evil'. We will see later why good and evil aren't relative or cultural, but evil, in either case,

here concerns suffering, which is not a man-made concept. Man-made evils are things like torture, murder, famine (yes, famines are man-made), and so on.

Second, we have **natural evil**. This is the kind of suffering which is not caused by man, but nature doing its thing. Since we are talking about suffering when we use the term 'evil', it should be clear that natural evils are a thing, because there are some sufferings caused by nature doing as it does. Here we have things like earthquakes, disease, volcanic eruptions, tornado, and sharknados.

The stronger of the two, for our purposes, concerns natural evils, because moral evils can be easily explained as not God's doing, but man's. But, unless I specify otherwise, when I just use the term 'evil' or 'evils', I am referring to the set of both. That being said, here is the Problem of Evil Argument:

1. There are evils in the world.
2. If God exists, He is all-knowing, all-powerful, and all-good.
3. If He is all knowing, then (a) He knows about the evil.
4. If He is all powerful, then (b) He can stop the evil.
5. If He is all good, then (c) He would want to stop the evil.
6. If (a), (b), and (c), then there would not be these evils in the world.
7. Therefore, God does not exist.

As you might guess, the most ripe for attack, the one which the theists really want to take down, is line 5. Typically, the problems or claimed solutions to POE which keep God in the mix, give some reason why God wouldn't want to stop the evil which is in the world. But, the vast majority of those counter points tend to fall flat.



## The Saints and Heroes Response

This is the claim that God allows for evil in the world because it leads to greater goods which could not have been had otherwise. Evils are necessary for the even greater good of having saints and heroes. When we learn about these evils we are called to arms against them, we are made heroes. Take for example this story, the events happened, and this only really works as an example of man-made evils:

In the First Crusade, there were several armies. Some of them were more official and organized than others. An unofficial, arranged and poorly managed group in the first crusade was lead by a Count Emicho from Germany. Emicho's plan for this group was for them to pass from Germany to the Holy Land immediately, which was a major issue because of food scarcity during that time of the year. Rather than following this ill-conceived plan, they eventually thought that the trek to the Holy Land was too far, and noticed that the local Jewish population was a lot closer and unarmed. As a result, they went from town to town, their own towns, pillaging and slaughtering the Jewish people in the region. The Church, to their credit, was vehemently opposed to this. They at the very least wanted the fighting outside of Christian dominated territory. Despite this horrific time, heroes arose to fight this evil. Priests and bishops not only preached against the violence, but even hid the Jewish population in their homes, fortifying the place. Some, if they had the means, raised personal armies to protect them. They used their religious might to protect Jewish places of worship and did their best to save lives.

If this suffering had not happened, we would not have the good of having these sorts of role models, people who would fight members of their own religion to protect people outside of

that circle. We would not have the good of these heroes being in the world.

An alternative version of this, which does escape some of the problem below is called the soul-making theodicy. This is that God allows for evil in the world to refine people, to make them better, a good which could not otherwise be gotten. Some philosophers really like this soul-making theodicy, stating that the purpose of evil is to allow free beings to grow and develop into those of the "finest characteristics". The Christian philosopher John Hick proposes this sort of reasoning in his work.<sup>109</sup> Other reasons like this are related: For example, God allows for evil as a test of faith and guides people towards God and stronger faith, again something which is said to not be able to get otherwise. This explanation for evil seems to be quite popular among classical Islamic philosophers.<sup>110</sup>

The problem with this is that much of the natural evils in the world are not recorded, we are not called to arms against them, and therefore could not have served to create the greater good. (Things like genetic diseases which lead to a slow and painful death are great examples, especially when there is no possible cure).

## The Artful Analogy

Some claim that the world is like God's canvas, He is painting a picture. There are shades of good everywhere and so too are there shades of evil. The contrasting elements of good and evil make the picture more beautiful than if it was only all shades of good. The goodness of the world is a product of the harmony between these shades of good and evil. Philosophers who have liked this kind of view, or at the very least wrote something

---

<sup>109</sup> John Hick, "The Problem of Evil," *The Encyclopedia of Philosophy*, edited by Paul Edwards (New York: Macmillan, 1967) p. 3.

<sup>110</sup> Nasrin Rouzati, "Evil and Human Suffering in Islamic Thought—Towards a Mystical Theodicy," *Religions* vol. 9, no. 2, 2018, <https://doi.org/10.3390/rel9020047>,

which is akin to it include: Leibniz in his *Theodicy*.<sup>111</sup> (paragraph number 213) and Augustine in his *Confessions* (book 7, chapters 13-16).<sup>112</sup> For Augustine, evil is there to make a more beautiful, harmonious whole of creation. We can't see the whole picture because we aren't God.

An apt analogy for this idea is to take two pictures (and for this analogy, the darker the shade, the more 'evil' it's meant to represent) and compare them. The first is just a blank white canvas, the other is a photo-realistic black-and-white picture of a rose. Which of these would you say is more beautiful? The rose or the blank canvas?

There are two major problems for this. The first is that it is hard to believe. Why would it be the case that a baby born with Tay-Sachs disease (a genetic disease which leads to a slow, painful, guaranteed death) would help the overall harmony of the world. Since God allows for so much evil and only God can see the picture, so to speak, then God must not be all-good in our sense, but in some other sense. True art is an explosion!

The second problem for this is that if God allows for the suffering to beautify the world, then God seems more like a sadist than an all-good entity. If suffering plays an artistic role in the world, then God is very close to a psychopathic artist who throws a bomb into a crowded elevator in order to admire the patterns created by the blast and the resultant splatter.

## The Free Will Defense

Philosophers danced around The Free Will Defense (FWD) as an explanation for Man-Made Evil for a lot of philosophic history, but was never really explicitly given. Most think that this is a very old and maybe obvious reply, but the current and, by far, most popular version, (even appearing in popular culture)

---

<sup>111</sup>Gottfried Wilhelm Leibniz, "Theodicy."

<sup>112</sup>Saint Augustine, *The Confessions* (Oxford UP UK, 1990).

is quite contemporary. The version which we will be discussing was given by Alvan Plantinga<sup>113</sup>

As a reply to the problem of evil, FWD is also extremely popular. This is the claim that human beings are special in that we have free will. If we lacked this aspect, we would be like robots or automata and could make no choices of our own. If you are familiar with Star Trek, if we lacked free will, then we would be like Data, but with emotions. Those that go with this defense claim that necessary consequence of having free will is the ability to do evil. Without this ability, then we would not really have free will. A world with free will and the possibility of doing evil is better than a world without free will. God made the best world possible, so, He made one with free will and thereby one with evil (suffering). The Free Will Defense goes like this, as before the bold lines are the premises:

### **1. Any world with free will is better than one without.**

This is the core intuition, or basis, for FWD. The thought here is that free will must be so valuable that any resulting suffering would be outweighed by it. Similar intuitions can be found in discussions about love/faith. For example, love is seen as a supreme good, but for a person to really love another, this must be voluntary, unforced, and originating, primarily, from the lover (or so it's claimed). This means that the person must have free will to really give love to another. A world with love is better than a loveless one. So, assuming that free will of this sort is required for love, a world with free will is better than one without.

---

<sup>113</sup>Alvin Plantinga, "The Free Will Defense," *Philosophy in America*, edited by Max Black (Ithaca: Cornell UP, 1965) pp. 204–20.

## **2. Any world with free will is going to have some evils (suffering) in it.**

This is a build-up from the previous line and will be attacked when we look at the problems with this defense. This is, sort of, I think, based on the intuition that with free will comes the actual ability to cause suffering and, as many of us have seen, if people have the ability to do something, someone, eventually will do it. Every aspect of this line, and the previous is suspect and very much open to rebuttal.

## **3. God made the best of all possible worlds.**

Since this is a reply to POE, we will see that it assumes the existence of God, which we will give as a pass. They are, in a sense, giving excuses for God's allowing evil. This is much like how a defense lawyer will assume that their client is innocent while in court. Since God is all-good, it makes sense that, if He made a world, it would be the best world possible for Him to make.

## **4. So, God made a world with free will.**

This follows from the first and third lines of the argument. Assuming that a world with free will is going to be better than one without it, it would follow that the best world which one could make would be one with free will. With that established, God making the best possible world would be a world with free will.

## **5. Therefore, God allows for evil as it's the result of free will.**

This follows from the second and fourth lines of the argument. Assuming, again, that a world with free will is going to have some suffering in it and that God made a world with free will, it follows that God must allow for the suffering in the world

because that suffering is a necessary consequence of having free will.

### **Connection with previous content:**

This should remind you about the stuff we have covered in Module 4. There, just as a refresher, we covered the various ideas and arguments for and against free-will. Be aware that the Free Will Defense requires incompatibilism and, more over, it requires libertarian free will. If you want to go with this defense, then you need to believe in Libertarianism. You must physically, not just counterfactually, be able to do evil and do otherwise than what you do. From here, we will continue down the free will defense rabbit hole and see some of the objections to this idea.

## **A Few Initial Replies to the Free Will Defense: How good is Free Will? Do we even have it?**

There are a few problems for the Free Will Defense, and they can be broken up into two different kinds, with a few exceptions. The first concerns the actual value of free will and the second concerns whether or not we actually have it.

### **Is Free Will that Good?**

This question relates to the first line of the argument, which claimed that any world with free will is going to be better than one without it. But, is that so obvious? The suffering in the world can be so great that it would make sense to want this possibility to go away. Here is a case to think about:

Imagine, if you will, two worlds, the first is a world with all of the suffering and anguish which is present in ours. In this world, people have this Libertarian Free Will. From a first person perspective, one could not tell the difference between 'making a choice' in this world vs in the world we presently live in. The second world is one without the vast majority of the suffering which we inflict on each other. For example, WW2 never happened in this world. This decrease in suffering is because there's no Libertarian Free Will in this world. Making a choice in this world, from a first person perspective, would feel exactly the same as making a choice in the world with free will. A person could not tell the difference. Which would you chose to live in? The one with unnecessary moral evils or the one without?

Similarly, we can there are certain moral evils which I will never get the chance to preform. For example, I will never get the opportunity to kick the pope in the shin. What if all evil acts were like that? What if people were never given the opportunity to do the evil acts or everyone had a moral character such that they would never do such things. Sure, they are possible, but we never get the chance to do them or would never do them. Am I less free because of my lacking the opportunity?

## Do We Actually Have it?

That last note leads us into an interesting other problem. Do we actually have free will? Or do we just have some kind of illusion of it? The idea of free will is hotly debated, as we saw in Module 4. Though this is a bit of a fallacy, the majority of philosophers out there believe that people have free will, but, this is not the kind of free will necessary for the Free Will Defense. Namely, most philosophers, at least the vast majority I know, think that compatibilism is true. In such a world, it is perfectly possible that we could have done otherwise (in the

counterfactual sense) but never actually do evil. More over, the choices we make, through various tests, can be shown to come from past experience and conditioning, we are programmed (so to speak). Which makes a serious problem for FWD.

## Free Will Without Evil

This one is related to the second kind of objection, namely, whether or not we have free will, but ties in to the second kind. One could argue that this objection gets the relevant parts of both. In this case we are asking whether it's possible to have free will without having the possibility to actually act/do evil. Much like the case which I gave before. The argument goes like this:

It is possible for all people to have free will and yet never bring about evil. God can bring about any possible situation. Therefore, God could bring about such a situation.

The issue here for FWD is that if this is possible, there's a better world than the one we have and therefore God didn't make the best one, taking all of the wind out of the sails of FWD. There have been some replies to this argument which are worth noting.

First, initially, people like to reject the first premise and claim that God could not have made a world like that. Saying something along the lines of "that would not really be free will!" However, going this route faces issues on its own. First, many claim that God is free but at the same time He/She does not bring about evil. So, as a result, why would it be impossible for man to be free and never bring about evil? Second, if we try to reject the first line, many of us think that God never wants us to commit evil acts. And yet, if we reject this line, it would mean that God is wishing for the impossible.

If you chose to reject the second line instead, you would be claiming that there are certain possible situations which God could not bring about. For example, it is possible for me to freely type this passage, but it's not possible for God to make it



the case that I freely type this passage because then it would not have been freely done; there's a contradiction. However, this particular route does not help in this case, it just shows that there are certain possible cases which God can't bring about. This, in turn, leads us down yet another rabbit hole about the very nature of being all-powerful. But, this also doesn't help because we can look at cases where people just aren't ever given the opportunity to act in certain ways. For example, it's a very serious possibility that I will never be given the opportunity to kick the pope in the shin. Although I would not want to do this, as a matter of fact, the fact remains that I will never be given the chance to do it. God could, quite easily, it would seem, make it the case that all opportunities to commit evil acts are like mine when it comes to kicking the pope in the shin, man is just never given the chance to do them. Would you say that I am less free because of this lack? If you say no, then we can have a world without evil and free will.

## Miracles

Another way in which God could get rid of, or at least limit, the number of evils in the world while retaining free will is to intervene more often.

People typically think that God intervenes in the world through miracles. But why does God do things which seem like minor tricks? Things like killing fig trees (one of Jesus' miracles), producing stigmata on people's hands, or making water into wine? Why doesn't He jump in and prevent or stop AIDS or the entirety of WW2? All God would need to do is whisper into people's ears at certain times to prevent them from committing the truly horrible acts of history.

The reply to this is typically that if God intervened then it would not be free will... But this doesn't really sit with the idea of miracles.

## The Problem of Natural Evil

The Free Will Defense, if it can surmount the problems which we have seen previously, would be able to handle the problem of evil, but only in a restricted sense. The Free Will Defense can only handle Man-Made Evil. But, there are other kinds of evils in the world. Remember, we also have Natural Evils to contend with. Remember, also, that whatever doesn't kill the Problem of Evil only makes it stronger. This leads us to the Problem of Natural Evil (or PONE, because it 'PONES' the Free Will Defense). There is no real connection (at least from the view of a non-believer) between free will and the plague, earthquakes, volcanic eruptions, etc. This argument is taking this into account. Rather than taking evils as any old kind of 'bad' in the world, here we will focus on natural evils, the pain and suffering caused by nature doing its thing.

1. There are natural evils in the world.
2. If God exists, He is all-knowing, all-powerful, and all-good.
3. If He is all knowing, then (a) He knows about the natural evils.
4. If He is all powerful, then (b) He can stop the natural evils.
5. If He is all good, then (c) He would want to stop the natural evils.
6. If (a), (b), and (c), then there would not be these natural evils in the world.
7. Therefore, God does not exist.

This should look very familiar from the regular old Problem of Evil. The Problem of Natural Evil is far stronger because it limits the scope. It is possible, however, to use some of the

responses to the Problem of Evil in response to the Problem of Natural Evil, but they are significantly weaker.

## **Saints and Heroes for The Problem of Natural Evil**

With the Saints and Heroes Response, we have that evil is in the world to make us better, to give us something to fight against. One way to think about this was that our souls are rough when we are created and the evils in the world are there to put us through the proverbial rock-tumbler. Natural Evils are there to give us something to fight against, having diseases gives us the goal of finding a cure and that process refines us.

But this does have its issues too, much like the regular old Saints and Heroes Response. For example, it seems plausible, and in fact, likely, that the moral evils in the world are enough to spur people into action. The evils which people do to each other constantly need to be fought against and the additional natural evils seem like an extra, unneeded, addition to get the saints and heroes. God would, if She is all good, want to refine us with as little suffering as possible.

## **The Artful Analogy for The Problem of Natural Evil**

For this reply, it was claimed that God allows for evil/suffering in the world because it makes the world more beautiful. God is all-good, in this sense, means that God is not morally perfect, but rather good as in a great artist. Following this line of thought, it's possible that the process of the environment restoring itself is beautiful in the eyes of God. Natural evils are the destruction necessary to rebuild and reshape.

This one has the same issues as before, but also one unique to it. For example, rather than resetting the already vibrant areas, why not just make another planet and start fresh? This would get the good of making something new without the suffering. Or in another case, when you have constructed something

beautiful in Minecraft and get bored of it, do you plant TNT around it and blow it up or do you save it and start building in a new world?

## The Fall Reply

Forgive me for my lack of familiarity with the creation story in Christianity and other Abrahamic religions. Some claim that the natural evils in the world are not God's doing, but are the doing of Adam and Eve in the Garden of Eden. This causes all natural evils to actually be man-made evils, which the Free Will Defense seems to be able to handle (maybe). In eating the apple, they created all of the natural evils in the world. We can think of the natural evils in the world as the punishment for Adam and Eve's initial breaking of the rule "don't eat from this tree." This sort of response is often found coming from the Biblical literalists, those who hold that the Earth was created in 7 days and all that.

There are a few issues with this reply, however. First, if God is all-good, then it would seem that He would need to be all-just (as in, never violating justice). Does it seem fair or correct, morally speaking, to punish the children for the actions of their parents? This seems especially wrong when the punished did not in anyway benefit from the transgression. For example, suppose that hundreds of years ago, my ancestor attempted to steal a pig. This pig ran away and back to the owner. The laws of the day say that even attempted pig stealing is punishable with the removing of the thumbs. Luckily, my ancestor was not caught. Generations down the line, it's discovered that my ancestor did this. Would it be morally permissible for them to punish me for the attempted theft, even though I had nothing to do with it? Is it OK to punish someone for the actions of their ancestor thousands of years ago?

Notice that I said "thousands of years ago". The second issue with this reply is that it requires that the story of Adam and Eve be an actual historical event. This can't be merely

mytho-historic (where we say we came from). There are multiple cases where this does not line up with science and other reliable aspects of the world. But, if those can be surmounted, then this would not be an issue.

## The Demon Reply

This reply is often cited to St. Augustine and it sort of has a feel like the phrase "the Devil made me do it". This follows from a similar line of thought as the Fall Reply, but also from a similar line of thought as the Free Will Defense. Here, like with the Fall Reply, they are trying to make a connection between these natural evils and 'moral evils' (evils by free agents (not necessarily human)). The claim is that God did not just give free will to humans, but also to certain divine beings (angels). This means that the actions of these divine entities are explained away as man-made evils, and God is not responsible for it. When Lucifer and the other Fallen rebelled, they did so freely. It's those Fallen angels which cause the 'natural' evils in the world. Since man's free will allows for us to do evil and it's still good, these demons are in the same sort of camp. In fact, many demons are claimed to be responsible for various kinds of natural disasters. Such as Furfur who is said to create tempests.

The responses to this sort of reply are varied. I encourage you to think of some and try them out in your heads.

## The Punishment Reply

Although this one, also, does not rely on the Fall, it's an option. This is the claim that natural disasters are the result of people committing sins, but not necessarily the sins of your ancestors. So, if enough people eat shrimp, God will send down a tidal wave. I thought, for a long time, that this sort of thinking was historical and that no one really believed it, however, this kind of thinking/explanation can be found in very extreme and offen-

sive sects of Christianity, such as the Westboro Baptist Church (don't look them up, they don't deserve the traffic). The reasoning here is that when a crime is committed, it's unjust for that crime to go unpunished. A just world is better than an unjust one, so a world with the kind of punishment dealt by natural evils is the better of possible options.

A quick and easy reply to this general idea runs down a similar line of reasoning as one we saw in the Fall Reply. Justice seems to require that the person who committed the crime be the one who receives the punishment. However, more often than not, these natural disasters seemingly hit indiscriminately. Those who are the most affected are often those who least deserve it. For example, the extremely poor.

### **The Good-Needs-Evil Reply**

This is a very classic claim. Essentially, although the nature of good and evil are objective, there is a relation between them. Good, in this sense, is the opposite of evil. Like 'left' and 'right', one can't logically have one without the other. So, God allows for evils because they are a logical consequence of having goods. One issue with this is that the nature of natural evils seems to be too much. Why have natural evils when moral evils will do?

### **The Beneficial Laws of Nature Defense**

Called BLOND for short, this stance claims that the laws of nature result in the disasters and other sufferings caused, but they have an outweighing good. This good is that they are consistent. Having consistent, regular laws of nature give rise to the processes which lead to humans being able to evolve and have free will. This could not have come about any other way (the law of nature needed to be as they are, this is an assumption worth examining), so the natural evils are just as much a byproduct of having beings with free will as the man-made evils. There are other aspects of this which could be seen as good, for

example, having the laws of nature be constant and unchanging allows for the good of scientific discovery and exploration. If there was some inconsistent aspect of these laws, eventually we would discover it and be unable to progress further.

But there are some issues for this stance. The first problem deals with the all-powerful nature of God. This does not explain why an all-powerful God chose those laws of nature. It is perfectly conceivable that He could have made laws which (in conjunction with other things) did not lead to natural evils. Similarly, God could have made it so that natural evils don't happen on planets where life would arise. Man-made evils, sure, but not natural ones. Intelligent beings, upon developing high enough, could discover the more naturally violent worlds, and thereby further marvel at the wonder of divine creation.

A reply to this is to say that even God is bound by the laws of nature. And, as a result, is not really all-powerful. But this also detracts from the assumptions made to get the defense started. In a related aspect, miracles are often pointed to as God's work, but those are violating the laws of nature, or at least bending them, so if God is bound by the Laws of Nature, how could they happen?

The second problem involves the idea of miracles. If we believe that God intervenes in the ways that He does, through miracles, then we need to ask why He engages with us in certain ways. Why turn water into wine when He could freeze the volcano? Every natural evil which are the byproduct of the laws of nature could be prevented by a being as powerful as God. If God never intervenes (as some claim, because that would hinder free will), then there would never have been any miracles.

## MODULE VI

*What is Knowledge?*

*How Do We Know*

*Things?*



# *Meditations on First Philosophy by René Descartes*

114

Translated by John Veitch (1901)

## **Meditation I OF THE THINGS OF WHICH WE MAY DOUBT.**

1. SEVERAL years have now elapsed since I first became aware that I had accepted, even from my youth, many false opinions for true, and that consequently what I afterward based on such principles was highly doubtful; and from that time I was convinced of the necessity of undertaking once in my life to rid myself of all the opinions I had adopted, and of commencing anew the work of building from the foundation, if I desired to establish a firm and abiding superstructure in the sciences. But

---

<sup>114</sup>René Descartes, *Meditations on First Philosophy* (Caravan Books, 1641).

as this enterprise appeared to me to be one of great magnitude, I waited until I had attained an age so mature as to leave me no hope that at any stage of life more advanced I should be better able to execute my design. On this account, I have delayed so long that I should henceforth consider I was doing wrong were I still to consume in deliberation any of the time that now remains for action. To-day, then, since I have opportunely freed my mind from all cares [and am happily disturbed by no passions], and since I am in the secure possession of leisure in a peaceable retirement, I will at length apply myself earnestly and freely to the general overthrow of all my former opinions.

2. But, to this end, it will not be necessary for me to show that the whole of these are false—a point, perhaps, which I shall never reach; but as even now my reason convinces me that I ought not the less carefully to withhold belief from what is not entirely certain and indubitable, than from what is manifestly false, it will be sufficient to justify the rejection of the whole if I shall find in each some ground for doubt. Nor for this purpose will it be necessary even to deal with each belief individually, which would be truly an endless labor; but, as the removal from below of the foundation necessarily involves the downfall of the whole edifice, I will at once approach the criticism of the principles on which all my former beliefs rested.

3. All that I have, up to this moment, accepted as possessed of the highest truth and certainty, I received either from or through the senses. I observed, however, that these sometimes misled us; and it is the part of prudence not to place absolute confidence in that by which we have even once been deceived.

4. But it may be said, perhaps, that, although the senses occasionally mislead us respecting minute objects, and such as are so far removed from us as to be beyond the reach of close observation, there are yet many other of their informations (presentations), of the truth of which it is manifestly impossible to doubt; as for example, that I am in this place, seated by the fire, clothed in a winter dressing gown, that I hold in my hands this piece of paper, with other intimations of the same nature.

But how could I deny that I possess these hands and this body, and withal escape being classed with persons in a state of insanity, whose brains are so disordered and clouded by dark bilious vapors as to cause them pertinaciously to assert that they are monarchs when they are in the greatest poverty; or clothed [in gold] and purple when destitute of any covering; or that their head is made of clay, their body of glass, or that they are gourds? I should certainly be not less insane than they, were I to regulate my procedure according to examples so extravagant.

5. Though this be true, I must nevertheless here consider that I am a man, and that, consequently, I am in the habit of sleeping, and representing to myself in dreams those same things, or even sometimes others less probable, which the insane think are presented to them in their waking moments. How often have I dreamt that I was in these familiar circumstances, that I was dressed, and occupied this place by the fire, when I was lying undressed in bed? At the present moment, however, I certainly look upon this paper with eyes wide awake; the head which I now move is not asleep; I extend this hand consciously and with express purpose, and I perceive it; the occurrences in sleep are not so distinct as all this. But I cannot forget that, at other times I have been deceived in sleep by similar illusions; and, attentively considering those cases, I perceive so clearly that there exist no certain marks by which the state of waking can ever be distinguished from sleep, that I feel greatly astonished; and in amazement I almost persuade myself that I am now dreaming.

6. Let us suppose, then, that we are dreaming, and that all these particulars—namely, the opening of the eyes, the motion of the head, the forth-putting of the hands—are merely illusions; and even that we really possess neither an entire body nor hands such as we see. Nevertheless it must be admitted at least that the objects which appear to us in sleep are, as it were, painted representations which could not have been formed unless in the likeness of realities; and, therefore, that those general objects, at all events, namely, eyes, a head, hands, and an en-

tire body, are not simply imaginary, but really existent. For, in truth, painters themselves, even when they study to represent sirens and satyrs by forms the most fantastic and extraordinary, cannot bestow upon them natures absolutely new, but can only make a certain medley of the members of different animals; or if they chance to imagine something so novel that nothing at all similar has ever been seen before, and such as is, therefore, purely fictitious and absolutely false, it is at least certain that the colors of which this is composed are real. And on the same principle, although these general objects, viz. [a body], eyes, a head, hands, and the like, be imaginary, we are nevertheless absolutely necessitated to admit the reality at least of some other objects still more simple and universal than these, of which, just as of certain real colors, all those images of things, whether true and real, or false and fantastic, that are found in our consciousness (*cogitatio*), are formed.

7. To this class of objects seem to belong corporeal nature in general and its extension; the figure of extended things, their quantity or magnitude, and their number, as also the place in, and the time during, which they exist, and other things of the same sort.

8. We will not, therefore, perhaps reason illegitimately if we conclude from this that Physics, Astronomy, Medicine, and all the other sciences that have for their end the consideration of composite objects, are indeed of a doubtful character; but that Arithmetic, Geometry, and the other sciences of the same class, which regard merely the simplest and most general objects, and scarcely inquire whether or not these are really existent, contain somewhat that is certain and indubitable: for whether I am awake or dreaming, it remains true that two and three make five, and that a square has but four sides; nor does it seem possible that truths so apparent can ever fall under a suspicion of falsity [or incertitude].

9. Nevertheless, the belief that there is a God who is all powerful, and who created me, such as I am, has, for a long time, obtained steady possession of my mind. How, then, do I know

that he has not arranged that there should be neither earth, nor sky, nor any extended thing, nor figure, nor magnitude, nor place, providing at the same time, however, for [the rise in me of the perceptions of all these objects, and] the persuasion that these do not exist otherwise than as I perceive them ? And further, as I sometimes think that others are in error respecting matters of which they believe themselves to possess a perfect knowledge, how do I know that I am not also deceived each time I add together two and three, or number the sides of a square, or form some judgment still more simple, if more simple indeed can be imagined? But perhaps Deity has not been willing that I should be thus deceived, for he is said to be supremely good. If, however, it were repugnant to the goodness of Deity to have created me subject to constant deception, it would seem likewise to be contrary to his goodness to allow me to be occasionally deceived; and yet it is clear that this is permitted.

10. Some, indeed, might perhaps be found who would be disposed rather to deny the existence of a Being so powerful than to believe that there is nothing certain. But let us for the present refrain from opposing this opinion, and grant that all which is here said of a Deity is fabulous: nevertheless, in whatever way it be supposed that I reach the state in which I exist, whether by fate, or chance, or by an endless series of antecedents and consequents, or by any other means, it is clear (since to be deceived and to err is a certain defect) that the probability of my being so imperfect as to be the constant victim of deception, will be increased exactly in proportion as the power possessed by the cause, to which they assign my origin, is lessened. To these reasonings I have assuredly nothing to reply, but am constrained at last to avow that there is nothing of all that I formerly believed to be true of which it is impossible to doubt, and that not through thoughtlessness or levity, but from cogent and maturely considered reasons; so that henceforward, if I desire to discover anything certain, I ought not the less carefully to refrain from assenting to those same opinions than to what might be shown to be manifestly false.

11. But it is not sufficient to have made these observations; care must be taken likewise to keep them in remembrance. For those old and customary opinions perpetually recur—long and familiar usage giving them the right of occupying my mind, even almost against my will, and subduing my belief; nor will I lose the habit of deferring to them and confiding in them so long as I shall consider them to be what in truth they are, viz, opinions to some extent doubtful, as I have already shown, but still highly probable, and such as it is much more reasonable to believe than deny. It is for this reason I am persuaded that I shall not be doing wrong, if, taking an opposite judgment of deliberate design, I become my own deceiver, by supposing, for a time, that all those opinions are entirely false and imaginary, until at length, having thus balanced my old by my new prejudices, my judgment shall no longer be turned aside by perverted usage from the path that may conduct to the perception of truth. For I am assured that, meanwhile, there will arise neither peril nor error from this course, and that I cannot for the present yield too much to distrust, since the end I now seek is not action but knowledge.

12. I will suppose, then, not that Deity, who is sovereignly good and the fountain of truth, but that some malignant demon, who is at once exceedingly potent and deceitful, has employed all his artifice to deceive me; I will suppose that the sky, the air, the earth, colors, figures, sounds, and all external things, are nothing better than the illusions of dreams, by means of which this being has laid snares for my credulity; I will consider myself as without hands, eyes, flesh, blood, or any of the senses, and as falsely believing that I am possessed of these; I will continue resolutely fixed in this belief, and if indeed by this means it be not in my power to arrive at the knowledge of truth, I shall at least do what is in my power, viz., [suspend my judgment], and guard with settled purpose against giving my assent to what is false, and being imposed upon by this deceiver, whatever be his power and artifice. But this undertaking is arduous, and a certain indolence insensibly leads me back to my ordinary course

of life; and just as the captive, who, perchance, was enjoying in his dreams an imaginary liberty, when he begins to suspect that it is but a vision, dreads awakening, and conspires with the agreeable illusions that the deception may be prolonged; so I, of my own accord, fall back into the train of my former beliefs, and fear to arouse myself from my slumber, lest the time of laborious wakefulness that would succeed this quiet rest, in place of bringing any light of day, should prove inadequate to dispel the darkness that will arise from the difficulties that have now been raised.

## **Meditation II OF THE NATURE OF THE HUMAN MIND; AND THAT IT IS MORE EASILY KNOWN THAN THE BODY.**

1. The Meditation of yesterday has filled my mind with so many doubts, that it is no longer in my power to forget them. Nor do I see, meanwhile, any principle on which they can be resolved; and, just as if I had fallen all of a sudden into very deep water, I am so greatly disconcerted as to be unable either to plant my feet firmly on the bottom or sustain myself by swimming on the surface. I will, nevertheless, make an effort, and try anew the same path on which I had entered yesterday, that is, proceed by casting aside all that admits of the slightest doubt, not less than if I had discovered it to be absolutely false; and I will continue always in this track until I shall find something that is certain, or at least, if I can do nothing more, until I shall know with certainty that there is nothing certain. Archimedes, that he might transport the entire globe from the place it occupied to another, demanded only a point that was firm and immovable; so, also, I shall be entitled to entertain the highest expectations, if I am fortunate enough to discover only one thing that is certain and indubitable.

2. I suppose, accordingly, that all the things which I see are false (fictitious); I believe that none of those objects which my fallacious memory represents ever existed; I suppose that I possess no senses; I believe that body, figure, extension, motion, and place are merely fictions of my mind. What is there, then, that can be esteemed true ? Perhaps this only, that there is absolutely nothing certain.

3. But how do I know that there is not something different altogether from the objects I have now enumerated, of which it is impossible to entertain the slightest doubt? Is there not a God, or some being, by whatever name I may designate him, who causes these thoughts to arise in my mind ? But why suppose such a being, for it may be I myself am capable of producing them? Am I, then, at least not something? But I before denied that I possessed senses or a body; I hesitate, however, for what follows from that? Am I so dependent on the body and the senses that without these I cannot exist? But I had the persuasion that there was absolutely nothing in the world, that there was no sky and no earth, neither minds nor bodies; was I not, therefore, at the same time, persuaded that I did not exist? Far from it; I assuredly existed, since I was persuaded. But there is I know not what being, who is possessed at once of the highest power and the deepest cunning, who is constantly employing all his ingenuity in deceiving me. Doubtless, then, I exist, since I am deceived; and, let him deceive me as he may, he can never bring it about that I am nothing, so long as I shall be conscious that I am something. So that it must, in fine, be maintained, all things being maturely and carefully considered, that this proposition (*pronunciatum*) I am, I exist, is necessarily true each time it is expressed by me, or conceived in my mind.

4. But I do not yet know with sufficient clearness what I am, though assured that I am; and hence, in the next place, I must take care, lest perchance I inconsiderately substitute some other object in room of what is properly myself, and thus wander from truth, even in that knowledge (*cognition*) which



I hold to be of all others the most certain and evident. For this reason, I will now consider anew what I formerly believed myself to be, before I entered on the present train of thought; and of my previous opinion I will retrench all that can in the least be invalidated by the grounds of doubt I have adduced, in order that there may at length remain nothing but what is certain and indubitable.

5. What then did I formerly think I was ? Undoubtedly I judged that I was a man. But what is a man ? Shall I say a rational animal ? Assuredly not; for it would be necessary forthwith to inquire into what is meant by animal, and what by rational, and thus, from a single question, I should insensibly glide into others, and these more difficult than the first; nor do I now possess enough of leisure to warrant me in wasting my time amid subtleties of this sort. I prefer here to attend to the thoughts that sprung up of themselves in my mind, and were inspired by my own nature alone, when I applied myself to the consideration of what I was. In the first place, then, I thought that I possessed a countenance, hands, arms, and all the fabric of members that appears in a corpse, and which I called by the name of body. It further occurred to me that I was nourished, that I walked, perceived, and thought, and all those actions I referred to the soul; but what the soul itself was I either did not stay to consider, or, if I did, I imagined that it was something extremely rare and subtile, like wind, or flame, or ether, spread through my grosser parts. As regarded the body, I did not even doubt of its nature, but thought I distinctly knew it, and if I had wished to describe it according to the notions I then entertained, I should have explained myself in this manner: By body I understand all that can be terminated by a certain figure; that can be comprised in a certain place, and so fill a certain space as therefrom to exclude every other body; that can be perceived either by touch, sight, hearing, taste, or smell; that can be moved in different ways, not indeed of itself, but by something foreign to it by which it is touched [and from which it receives the impression]; for the power of self-

motion, as likewise that of perceiving and thinking, I held as by no means pertaining to the nature of body; on the contrary, I was somewhat astonished to find such faculties existing in some bodies.

6. But [as to myself, what can I now say that I am], since I suppose there exists an extremely powerful, and, if I may so speak, malignant being, whose whole endeavors are directed toward deceiving me ? Can I affirm that I possess any one of all those attributes of which I have lately spoken as belonging to the nature of body ? After attentively considering them in my own mind, I find none of them that can properly be said to belong to myself. To recount them were idle and tedious. Let us pass, then, to the attributes of the soul. The first mentioned were the powers of nutrition and walking; but, if it be true that I have no body, it is true likewise that I am capable neither of walking nor of being nourished. Perception is another attribute of the soul; but perception too is impossible without the body; besides, I have frequently, during sleep, believed that I perceived objects which I afterward observed I did not in reality perceive. Thinking is another attribute of the soul; and here I discover what properly belongs to myself. This alone is inseparable from me. I am—I exist: this is certain; but how often? As often as I think; for perhaps it would even happen, if I should wholly cease to think, that I should at the same time altogether cease to be. I now admit nothing that is not necessarily true. I am therefore, precisely speaking, only a thinking thing, that is, a mind (*mens sive animus*), understanding, or reason, terms whose signification was before unknown to me. I am, however, a real thing, and really existent; but what thing? The answer was, a thinking thing.

7. The question now arises, am I aught besides ? I will stimulate my imagination with a view to discover whether I am not still something more than a thinking being. Now it is plain I am not the assemblage of members called the human body; I am not a thin and penetrating air diffused through all these members, or wind, or flame, or vapor, or breath, or any of all

the things I can imagine; for I supposed that all these were not, and, without changing the supposition, I find that I still feel assured of my existence. But it is true, perhaps, that those very things which I suppose to be non-existent, because they are unknown to me, are not in truth different from myself whom I know. This is a point I cannot determine, and do not now enter into any dispute regarding it. I can only judge of things that are known to me: I am conscious that I exist, and I who know that I exist inquire into what I am. It is, however, perfectly certain that the knowledge of my existence, thus precisely taken, is not dependent on things, the existence of which is as yet unknown to me: and consequently it is not dependent on any of the things I can feign in imagination. Moreover, the phrase itself, I frame an image (*effingo*), reminds me of my error; for I should in truth frame one if I were to imagine myself to be anything, since to imagine is nothing more than to contemplate the figure or image of a corporeal thing; but I already know that I exist, and that it is possible at the same time that all those images, and in general all that relates to the nature of body, are merely dreams [or *chimeras*]. From this I discover that it is not more reasonable to say, I will excite my imagination that I may know more distinctly what I am, than to express myself as follows: I am now awake, and perceive something real; but because my perception is not sufficiently clear, I will of express purpose go to sleep that my dreams may represent to me the object of my perception with more truth and clearness. And, therefore, I know that nothing of all that I can embrace in imagination belongs to the knowledge which I have of myself, and that there is need to recall with the utmost care the mind from this mode of thinking, that it may be able to know its own nature with perfect distinctness.

8. But what, then, am I ? A thinking thing, it has been said. But what is a thinking thing? It is a thing that doubts, understands, [conceives], affirms, denies, wills, refuses; that imagines also, and perceives.

9. Assuredly it is not little, if all these properties belong to

my nature. But why should they not belong to it ? Am I not that very being who now doubts of almost everything; who, for all that, understands and conceives certain things; who affirms one alone as true, and denies the others; who desires to know more of them, and does not wish to be deceived; who imagines many things, sometimes even despite his will; and is likewise percipient of many, as if through the medium of the senses. Is there nothing of all this as true as that I am, even although I should be always dreaming, and although he who gave me being employed all his ingenuity to deceive me ? Is there also any one of these attributes that can be properly distinguished from my thought, or that can be said to be separate from myself ? For it is of itself so evident that it is I who doubt, I who understand, and I who desire, that it is here unnecessary to add anything by way of rendering it more clear. And I am as certainly the same being who imagines; for although it may be (as I before supposed) that nothing I imagine is true, still the power of imagination does not cease really to exist in me and to form part of my thought. In fine, I am the same being who perceives, that is, who apprehends certain objects as by the organs of sense, since, in truth, I see light, hear a noise, and feel heat. But it will be said that these presentations are false, and that I am dreaming. Let it be so. At all events it is certain that I seem to see light, hear a noise, and feel heat; this cannot be false, and this is what in me is properly called perceiving (sentire), which is nothing else than thinking.

10. From this I begin to know what I am with somewhat greater clearness and distinctness than heretofore. But, nevertheless, it still seems to me, and I cannot help believing, that corporeal things, whose images are formed by thought [which fall under the senses], and are examined by the same, are known with much greater distinctness than that I know not what part of myself which is not imaginable; although, in truth, it may seem strange to say that I know and comprehend with greater distinctness things whose existence appears to me doubtful, that are unknown, and do not belong to me, than others of

whose reality I am persuaded, that are known to me, and appertain to my proper nature; in a word, than myself. But I see clearly what is the state of the case. My mind is apt to wander, and will not yet submit to be restrained within the limits of truth. Let us therefore leave the mind to itself once more, and, according to it every kind of liberty [permit it to consider the objects that appear to it from without], in order that, having afterward withdrawn it from these gently and opportunely [and fixed it on the consideration of its being and the properties it finds in itself], it may then be the more easily controlled.

11. Let us now accordingly consider the objects that are commonly thought to be [the most easily, and likewise] the most distinctly known, viz, the bodies we touch and see; not, indeed, bodies in general, for these general notions are usually somewhat more confused, but one body in particular. Take, for example, this piece of wax; it is quite fresh, having been but recently taken from the beehive; it has not yet lost the sweetness of the honey it contained; it still retains somewhat of the odor of the flowers from which it was gathered; its color, figure, size, are apparent (to the sight); it is hard, cold, easily handled; and sounds when struck upon with the finger. In fine, all that contributes to make a body as distinctly known as possible, is found in the one before us. But, while I am speaking, let it be placed near the fire—what remained of the taste exhales, the smell evaporates, the color changes, its figure is destroyed, its size increases, it becomes liquid, it grows hot, it can hardly be handled, and, although struck upon, it emits no sound. Does the same wax still remain after this change? It must be admitted that it does remain; no one doubts it, or judges otherwise. What, then, was it I knew with so much distinctness in the piece of wax? Assuredly, it could be nothing of all that I observed by means of the senses, since all the things that fell under taste, smell, sight, touch, and hearing are changed, and yet the same wax remains.

12. It was perhaps what I now think, viz, that this wax was neither the sweetness of honey, the pleasant odor of flow-

ers, the whiteness, the figure, nor the sound, but only a body that a little before appeared to me conspicuous under these forms, and which is now perceived under others. But, to speak precisely, what is it that I imagine when I think of it in this way? Let it be attentively considered, and, retrenching all that does not belong to the wax, let us see what remains. There certainly remains nothing, except something extended, flexible, and movable. But what is meant by flexible and movable? Is it not that I imagine that the piece of wax, being round, is capable of becoming square, or of passing from a square into a triangular figure? Assuredly such is not the case, because I conceive that it admits of an infinity of similar changes; and I am, moreover, unable to compass this infinity by imagination, and consequently this conception which I have of the wax is not the product of the faculty of imagination. But what now is this extension? Is it not also unknown? for it becomes greater when the wax is melted, greater when it is boiled, and greater still when the heat increases; and I should not conceive [clearly and] according to truth, the wax as it is, if I did not suppose that the piece we are considering admitted even of a wider variety of extension than I ever imagined, I must, therefore, admit that I cannot even comprehend by imagination what the piece of wax is, and that it is the mind alone (*mens*, Lat., *entendement*, F.) which perceives it. I speak of one piece in particular; for as to wax in general, this is still more evident. But what is the piece of wax that can be perceived only by the [understanding or] mind? It is certainly the same which I see, touch, imagine; and, in fine, it is the same which, from the beginning, I believed it to be. But (and this it is of moment to observe) the perception of it is neither an act of sight, of touch, nor of imagination, and never was either of these, though it might formerly seem so, but is simply an intuition (*inspectio*) of the mind, which may be imperfect and confused, as it formerly was, or very clear and distinct, as it is at present, according as the attention is more or less directed to the elements which it contains, and of which it is composed.

13. But, meanwhile, I feel greatly astonished when I observe [the weakness of my mind, and] its proneness to error. For although, without at all giving expression to what I think, I consider all this in my own mind, words yet occasionally impede my progress, and I am almost led into error by the terms of ordinary language. We say, for example, that we see the same wax when it is before us, and not that we judge it to be the same from its retaining the same color and figure: whence I should forthwith be disposed to conclude that the wax is known by the act of sight, and not by the intuition of the mind alone, were it not for the analogous instance of human beings passing on in the street below, as observed from a window. In this case I do not fail to say that I see the men themselves, just as I say that I see the wax; and yet what do I see from the window beyond hats and cloaks that might cover artificial machines, whose motions might be determined by springs ? But I judge that there are human beings from these appearances, and thus I comprehend, by the faculty of judgment alone which is in the mind, what I believed I saw with my eyes.

14. The man who makes it his aim to rise to knowledge superior to the common, ought to be ashamed to seek occasions of doubting from the vulgar forms of speech: instead, therefore, of doing this, I shall proceed with the matter in hand, and inquire whether I had a clearer and more perfect perception of the piece of wax when I first saw it, and when I thought I knew it by means of the external sense itself, or, at all events, by the common sense (*sensus communis*), as it is called, that is, by the imaginative faculty; or whether I rather apprehend it more clearly at present, after having examined with greater care, both what it is, and in what way it can be known. It would certainly be ridiculous to entertain any doubt on this point. For what, in that first perception, was there distinct ? What did I perceive which any animal might not have perceived ? But when I distinguish the wax from its exterior forms, and when, as if I had stripped it of its vestments, I consider it quite naked, it is certain, although some error may still be found in

my judgment, that I cannot, nevertheless, thus apprehend it without possessing a human mind.

15. But finally, what shall I say of the mind itself, that is, of myself ? for as yet I do not admit that I am anything but mind. What, then! I who seem to possess so distinct an apprehension of the piece of wax, do I not know myself, both with greater truth and certitude, and also much more distinctly and clearly? For if I judge that the wax exists because I see it, it assuredly follows, much more evidently, that I myself am or exist, for the same reason: for it is possible that what I see may not in truth be wax, and that I do not even possess eyes with which to see anything; but it cannot be that when I see, or, which comes to the same thing, when I think I see, I myself who think am nothing. So likewise, if I judge that the wax exists because I touch it, it will still also follow that I am; and if I determine that my imagination, or any other cause, whatever it be, persuades me of the existence of the wax, I will still draw the same conclusion. And what is here remarked of the piece of wax, is applicable to all the other things that are external to me. And further, if the [notion or] perception of wax appeared to me more precise and distinct, after that not only sight and touch, but many other causes besides, rendered it manifest to my apprehension, with how much greater distinctness must I now know myself, since all the reasons that contribute to the knowledge of the nature of wax, or of any body whatever, manifest still better the nature of my mind ? And there are besides so many other things in the mind itself that contribute to the illustration of its nature, that those dependent on the body, to which I have here referred, scarcely merit to be taken into account.

16. But, in conclusion, I find I have insensibly reverted to the point I desired; for, since it is now manifest to me that bodies themselves are not properly perceived by the senses nor by the faculty of imagination, but by the intellect alone; and since they are not perceived because they are seen and touched, but only because they are understood [or rightly comprehended



by thought], I readily discover that there is nothing more easily or clearly apprehended than my own mind. But because it is difficult to rid one's self so promptly of an opinion to which one has been long accustomed, it will be desirable to tarry for some time at this stage, that, by long continued meditation, I may more deeply impress upon my memory this new knowledge.

## **Meditation III OF GOD: THAT HE EXISTS.**

1. I WILL now close my eyes, I will stop my ears, I will turn away my senses from their objects, I will even efface from my consciousness all the images of corporeal things; or at least, because this can hardly be accomplished, I will consider them as empty and false; and thus, holding converse only with myself, and closely examining my nature, I will endeavor to obtain by degrees a more intimate and familiar knowledge of myself. I am a thinking (conscious) thing, that is, a being who doubts, affirms, denies, knows a few objects, and is ignorant of many,—[who loves, hates], wills, refuses, who imagines likewise, and perceives; for, as I before remarked, although the things which I perceive or imagine are perhaps nothing at all apart from me [and in themselves], I am nevertheless assured that those modes of consciousness which I call perceptions and imaginations, in as far only as they are modes of consciousness, exist in me.

2. And in the little I have said I think I have summed up all that I really know, or at least all that up to this time I was aware I knew. Now, as I am endeavoring to extend my knowledge more widely, I will use circumspection, and consider with care whether I can still discover in myself anything further which I have not yet hitherto observed. I am certain that I am a thinking thing; but do I not therefore likewise know what is required to render me certain of a truth? In this first knowledge, doubtless, there is nothing that gives me assurance of its truth except the clear and distinct perception of what I affirm,

which would not indeed be sufficient to give me the assurance that what I say is true, if it could ever happen that anything I thus clearly and distinctly perceived should prove false; and accordingly it seems to me that I may now take as a general rule, that all that is very clearly and distinctly apprehended (conceived) is true.

3. Nevertheless I before received and admitted many things as wholly certain and manifest, which yet I afterward found to be doubtful. What, then, were those? They were the earth, the sky, the stars, and all the other objects which I was in the habit of perceiving by the senses. But what was it that I clearly [and distinctly] perceived in them? Nothing more than that the ideas and the thoughts of those objects were presented to my mind. And even now I do not deny that these ideas are found in my mind. But there was yet another thing which I affirmed, and which, from having been accustomed to believe it, I thought I clearly perceived, although, in truth, I did not perceive it at all; I mean the existence of objects external to me, from which those ideas proceeded, and to which they had a perfect resemblance; and it was here I was mistaken, or if I judged correctly, this assuredly was not to be traced to any knowledge I possessed (the force of my perception, *Lat.*).

4. But when I considered any matter in arithmetic and geometry, that was very simple and easy, as, for example, that two and three added together make five, and things of this sort, did I not view them with at least sufficient clearness to warrant me in affirming their truth? Indeed, if I afterward judged that we ought to doubt of these things, it was for no other reason than because it occurred to me that God might perhaps have given me such a nature as that I should be deceived, even respecting the matters that appeared to me the most evidently true. But as often as this preconceived opinion of the sovereign power of a God presents itself to my mind, I am constrained to admit that it is easy for him, if he wishes it, to cause me to err, even in matters where I think I possess the highest evidence; and, on the other hand, as often as I direct my attention to

things which I think I apprehend with great clearness, I am so persuaded of their truth that I naturally break out into expressions such as these: Deceive me who may, no one will yet ever be able to bring it about that I am not, so long as I shall be conscious that I am, or at any future time cause it to be true that I have never been, it being now true that I am, or make two and three more or less than five, in supposing which, and other like absurdities, I discover a manifest contradiction. And in truth, as I have no ground for believing that Deity is deceitful, and as, indeed, I have not even considered the reasons by which the existence of a Deity of any kind is established, the ground of doubt that rests only on this supposition is very slight, and, so to speak, metaphysical. But, that I may be able wholly to remove it, I must inquire whether there is a God, as soon as an opportunity of doing so shall present itself; and if I find that there is a God, I must examine likewise whether he can be a deceiver; for, without the knowledge of these two truths, I do not see that I can ever be certain of anything. And that I may be enabled to examine this without interrupting the order of meditation I have proposed to myself [which is, to pass by degrees from the notions that I shall find first in my mind to those I shall afterward discover in it], it is necessary at this stage to divide all my thoughts into certain classes, and to consider in which of these classes truth and error are, strictly speaking, to be found.

5. Of my thoughts some are, as it were, images of things, and to these alone properly belongs the name *IDEA*; as when I think [represent to my mind] a man, a chimera, the sky, an angel or God. Others, again, have certain other forms; as when I will, fear, affirm, or deny, I always, indeed, apprehend something as the object of my thought, but I also embrace in thought something more than the representation of the object; and of this class of thoughts some are called volitions or affections, and others judgments.

6. Now, with respect to ideas, if these are considered only in themselves, and are not referred to any object beyond them,

they cannot, properly speaking, be false; for, whether I imagine a goat or chimera, it is not less true that I imagine the one than the other. Nor need we fear that falsity may exist in the will or affections; for, although I may desire objects that are wrong, and even that never existed, it is still true that I desire them. There thus only remain our judgments, in which we must take diligent heed that we be not deceived. But the chief and most ordinary error that arises in them consists in judging that the ideas which are in us are like or conformed to the things that are external to us; for assuredly, if we but considered the ideas themselves as certain modes of our thought (consciousness), without referring them to anything beyond, they would hardly afford any occasion of error.

7. But among these ideas, some appear to me to be innate, others adventitious, and others to be made by myself (factitious); for, as I have the power of conceiving what is called a thing, or a truth, or a thought, it seems to me that I hold this power from no other source than my own nature; but if I now hear a noise, if I see the sun, or if I feel heat, I have all along judged that these sensations proceeded from certain objects existing out of myself; and, in fine, it appears to me that sirens, hippogryphs, and the like, are inventions of my own mind. But I may even perhaps come to be of opinion that all my ideas are of the class which I call adventitious, or that they are all innate, or that they are all factitious; for I have not yet clearly discovered their true origin.

8. What I have here principally to do is to consider, with reference to those that appear to come from certain objects without me, what grounds there are for thinking them like these objects. The first of these grounds is that it seems to me I am so taught by nature; and the second that I am conscious that those ideas are not dependent on my will, and therefore not on myself, for they are frequently presented to me against my will, as at present, whether I will or not, I feel heat; and I am thus persuaded that this sensation or idea (*sensum vel ideam*) of heat is produced in me by something different from myself, viz., by

the heat of the fire by which I sit. And it is very reasonable to suppose that this object impresses me with its own likeness rather than any other thing.

9. But I must consider whether these reasons are sufficiently strong and convincing. When I speak of being taught by nature in this matter, I understand by the word nature only a certain spontaneous impetus that impels me to believe in a resemblance between ideas and their objects, and not a natural light that affords a knowledge of its truth. But these two things are widely different; for what the natural light shows to be true can be in no degree doubtful, as, for example, that I am because I doubt, and other truths of the like kind; inasmuch as I possess no other faculty whereby to distinguish truth from error, which can teach me the falsity of what the natural light declares to be true, and which is equally trustworthy; but with respect to [seemingly] natural impulses, I have observed, when the question related to the choice of right or wrong in action, that they frequently led me to take the worse part; nor do I see that I have any better ground for following them in what relates to truth and error.

10. Then, with respect to the other reason, which is that because these ideas do not depend on my will, they must arise from objects existing without me, I do not find it more convincing than the former, for just as those natural impulses, of which I have lately spoken, are found in me, notwithstanding that they are not always in harmony with my will, so likewise it may be that I possess some power not sufficiently known to myself capable of producing ideas without the aid of external objects, and, indeed, it has always hitherto appeared to me that they are formed during sleep, by some power of this nature, without the aid of aught external.

11. And, in fine, although I should grant that they proceeded from those objects, it is not a necessary consequence that they must be like them. On the contrary, I have observed, in a number of instances, that there was a great difference between the object and its idea. Thus, for example, I find in my mind two wholly diverse ideas of the sun; the one, by which it

appears to me extremely small draws its origin from the senses, and should be placed in the class of adventitious ideas; the other, by which it seems to be many times larger than the whole earth, is taken up on astronomical grounds, that is, elicited from certain notions born with me, or is framed by myself in some other manner. These two ideas cannot certainly both resemble the same sun; and reason teaches me that the one which seems to have immediately emanated from it is the most unlike.

12. And these things sufficiently prove that hitherto it has not been from a certain and deliberate judgment, but only from a sort of blind impulse, that I believed existence of certain things different from myself, which, by the organs of sense, or by whatever other means it might be, conveyed their ideas or images into my mind [and impressed it with their likenesses].

13. But there is still another way of inquiring whether, of the objects whose ideas are in my mind, there are any that exist out of me. If ideas are taken in so far only as they are certain modes of consciousness, I do not remark any difference or inequality among them, and all seem, in the same manner, to proceed from myself; but, considering them as images, of which one represents one thing and another a different, it is evident that a great diversity obtains among them. For, without doubt, those that represent substances are something more, and contain in themselves, so to speak, more objective reality [that is, participate by representation in higher degrees of being or perfection], than those that represent only modes or accidents; and again, the idea by which I conceive a God [sovereign], eternal, infinite, [immutable], all-knowing, all-powerful, and the creator of all things that are out of himself, this, I say, has certainly in it more objective reality than those ideas by which finite substances are represented.

14. Now, it is manifest by the natural light that there must at least be as much reality in the efficient and total cause as in its effect; for whence can the effect draw its reality if not from its cause? And how could the cause communicate to it this reality unless it possessed it in itself? And hence it follows, not only

that what is cannot be produced by what is not, but likewise that the more perfect, in other words, that which contains in itself more reality, cannot be the effect of the less perfect; and this is not only evidently true of those effects, whose reality is actual or formal, but likewise of ideas, whose reality is only considered as objective. Thus, for example, the stone that is not yet in existence, not only cannot now commence to be, unless it be produced by that which possesses in itself, formally or eminently, all that enters into its composition, [in other words, by that which contains in itself the same properties that are in the stone, or others superior to them]; and heat can only be produced in a subject that was before devoid of it, by a cause that is of an order, [degree or kind], at least as perfect as heat; and so of the others. But further, even the idea of the heat, or of the stone, cannot exist in me unless it be put there by a cause that contains, at least, as much reality as I conceive existent in the heat or in the stone for although that cause may not transmit into my idea anything of its actual or formal reality, we ought not on this account to imagine that it is less real; but we ought to consider that, [as every idea is a work of the mind], its nature is such as of itself to demand no other formal reality than that which it borrows from our consciousness, of which it is but a mode [that is, a manner or way of thinking]. But in order that an idea may contain this objective reality rather than that, it must doubtless derive it from some cause in which is found at least as much formal reality as the idea contains of objective; for, if we suppose that there is found in an idea anything which was not in its cause, it must of course derive this from nothing. But, however imperfect may be the mode of existence by which a thing is objectively [or by representation] in the understanding by its idea, we certainly cannot, for all that, allege that this mode of existence is nothing, nor, consequently, that the idea owes its origin to nothing.

15. Nor must it be imagined that, since the reality which considered in these ideas is only objective, the same reality need not be formally (actually) in the causes of these ideas, but

only objectively: for, just as the mode of existing objectively belongs to ideas by their peculiar nature, so likewise the mode of existing formally appertains to the causes of these ideas (at least to the first and principal), by their peculiar nature. And although an idea may give rise to another idea, this regress cannot, nevertheless, be infinite; we must in the end reach a first idea, the cause of which is, as it were, the archetype in which all the reality [or perfection] that is found objectively [or by representation] in these ideas is contained formally [and in act]. I am thus clearly taught by the natural light that ideas exist in me as pictures or images, which may, in truth, readily fall short of the perfection of the objects from which they are taken, but can never contain anything greater or more perfect.

16. And in proportion to the time and care with which I examine all those matters, the conviction of their truth brightens and becomes distinct. But, to sum up, what conclusion shall I draw from it all? It is this: if the objective reality [or perfection] of any one of my ideas be such as clearly to convince me, that this same reality exists in me neither formally nor eminently, and if, as follows from this, I myself cannot be the cause of it, it is a necessary consequence that I am not alone in the world, but that there is besides myself some other being who exists as the cause of that idea; while, on the contrary, if no such idea be found in my mind, I shall have no sufficient ground of assurance of the existence of any other being besides myself, for, after a most careful search, I have, up to this moment, been unable to discover any other ground.

17. But, among these my ideas, besides that which represents myself, respecting which there can be here no difficulty, there is one that represents a God; others that represent corporeal and inanimate things; others angels; others animals; and, finally, there are some that represent men like myself.

18. But with respect to the ideas that represent other men, or animals, or angels, I can easily suppose that they were formed by the mingling and composition of the other ideas which I have of myself, of corporeal things, and of God, although they were,



apart from myself, neither men, animals, nor angels.

19. And with regard to the ideas of corporeal objects, I never discovered in them anything so great or excellent which I myself did not appear capable of originating; for, by considering these ideas closely and scrutinizing them individually, in the same way that I yesterday examined the idea of wax, I find that there is but little in them that is clearly and distinctly perceived. As belonging to the class of things that are clearly apprehended, I recognize the following, viz, magnitude or extension in length, breadth, and depth; figure, which results from the termination of extension; situation, which bodies of diverse figures preserve with reference to each other; and motion or the change of situation; to which may be added substance, duration, and number. But with regard to light, colors, sounds, odors, tastes, heat, cold, and the other tactile qualities, they are thought with so much obscurity and confusion, that I cannot determine even whether they are true or false; in other words, whether or not the ideas I have of these qualities are in truth the ideas of real objects. For although I before remarked that it is only in judgments that formal falsity, or falsity properly so called, can be met with, there may nevertheless be found in ideas a certain material falsity, which arises when they represent what is nothing as if it were something. Thus, for example, the ideas I have of cold and heat are so far from being clear and distinct, that I am unable from them to discover whether cold is only the privation of heat, or heat the privation of cold; or whether they are or are not real qualities: and since, ideas being as it were images there can be none that does not seem to us to represent some object, the idea which represents cold as something real and positive will not improperly be called false, if it be correct to say that cold is nothing but a privation of heat; and so in other cases.

20. To ideas of this kind, indeed, it is not necessary that I should assign any author besides myself: for if they are false, that is, represent objects that are unreal, the natural light teaches me that they proceed from nothing; in other words,

that they are in me only because something is wanting to the perfection of my nature; but if these ideas are true, yet because they exhibit to me so little reality that I cannot even distinguish the object represented from nonbeing, I do not see why I should not be the author of them.

21. With reference to those ideas of corporeal things that are clear and distinct, there are some which, as appears to me, might have been taken from the idea I have of myself, as those of substance, duration, number, and the like. For when I think that a stone is a substance, or a thing capable of existing of itself, and that I am likewise a substance, although I conceive that I am a thinking and non-extended thing, and that the stone, on the contrary, is extended and unconscious, there being thus the greatest diversity between the two concepts, yet these two ideas seem to have this in common that they both represent substances. In the same way, when I think of myself as now existing, and recollect besides that I existed some time ago, and when I am conscious of various thoughts whose number I know, I then acquire the ideas of duration and number, which I can afterward transfer to as many objects as I please. With respect to the other qualities that go to make up the ideas of corporeal objects, viz, extension, figure, situation, and motion, it is true that they are not formally in me, since I am merely a thinking being; but because they are only certain modes of substance, and because I myself am a substance, it seems possible that they may be contained in me eminently.

22. There only remains, therefore, the idea of God, in which I must consider whether there is anything that cannot be supposed to originate with myself. By the name God, I understand a substance infinite, [eternal, immutable], independent, all-knowing, all-powerful, and by which I myself, and every other thing that exists, if any such there be, were created. But these properties are so great and excellent, that the more attentively I consider them the less I feel persuaded that the idea I have of them owes its origin to myself alone. And thus it is absolutely necessary to conclude, from all that I have before

said, that God exists.

23. For though the idea of substance be in my mind owing to this, that I myself am a substance, I should not, however, have the idea of an infinite substance, seeing I am a finite being, unless it were given me by some substance in reality infinite.

24. And I must not imagine that I do not apprehend the infinite by a true idea, but only by the negation of the finite, in the same way that I comprehend repose and darkness by the negation of motion and light: since, on the contrary, I clearly perceive that there is more reality in the infinite substance than in the finite, and therefore that in some way I possess the perception (notion) of the infinite before that of the finite, that is, the perception of God before that of myself, for how could I know that I doubt, desire, or that something is wanting to me, and that I am not wholly perfect, if I possessed no idea of a being more perfect than myself, by comparison of which I knew the deficiencies of my nature?

25. And it cannot be said that this idea of God is perhaps materially false, and consequently that it may have arisen from nothing [in other words, that it may exist in me from my imperfections as I before said of the ideas of heat and cold, and the like: for, on the contrary, as this idea is very clear and distinct, and contains in itself more objective reality than any other, there can be no one of itself more true, or less open to the suspicion of falsity. The idea, I say, of a being supremely perfect, and infinite, is in the highest degree true; for although, perhaps, we may imagine that such a being does not exist, we cannot, nevertheless, suppose that his idea represents nothing real, as I have already said of the idea of cold. It is likewise clear and distinct in the highest degree, since whatever the mind clearly and distinctly conceives as real or true, and as implying any perfection, is contained entire in this idea. And this is true, nevertheless, although I do not comprehend the infinite, and although there may be in God an infinity of things that I cannot comprehend, nor perhaps even compass by thought in any way; for it is of the nature of the infinite that it should not be

comprehended by the finite; and it is enough that I rightly understand this, and judge that all which I clearly perceive, and in which I know there is some perfection, and perhaps also an infinity of properties of which I am ignorant, are formally or eminently in God, in order that the idea I have of him may be come the most true, clear, and distinct of all the ideas in my mind.

26. But perhaps I am something more than I suppose myself to be, and it may be that all those perfections which I attribute to God, in some way exist potentially in me, although they do not yet show themselves, and are not reduced to act. Indeed, I am already conscious that my knowledge is being increased [and perfected] by degrees; and I see nothing to prevent it from thus gradually increasing to infinity, nor any reason why, after such increase and perfection, I should not be able thereby to acquire all the other perfections of the Divine nature; nor, in fine, why the power I possess of acquiring those perfections, if it really now exist in me, should not be sufficient to produce the ideas of them.

27. Yet, on looking more closely into the matter, I discover that this cannot be; for, in the first place, although it were true that my knowledge daily acquired new degrees of perfection, and although there were potentially in my nature much that was not as yet actually in it, still all these excellences make not the slightest approach to the idea I have of the Deity, in whom there is no perfection merely potentially [but all actually] existent; for it is even an unmistakable token of imperfection in my knowledge, that it is augmented by degrees. Further, although my knowledge increase more and more, nevertheless I am not, therefore, induced to think that it will ever be actually infinite, since it can never reach that point beyond which it shall be incapable of further increase. But I conceive God as actually infinite, so that nothing can be added to his perfection. And, in fine, I readily perceive that the objective being of an idea cannot be produced by a being that is merely potentially existent, which, properly speaking, is nothing, but only by a

being existing formally or actually.

28. And, truly, I see nothing in all that I have now said which it is not easy for any one, who shall carefully consider it, to discern by the natural light; but when I allow my attention in some degree to relax, the vision of my mind being obscured, and, as it were, blinded by the images of sensible objects, I do not readily remember the reason why the idea of a being more perfect than myself, must of necessity have proceeded from a being in reality more perfect. On this account I am here desirous to inquire further, whether I, who possess this idea of God, could exist supposing there were no God.

29. And I ask, from whom could I, in that case, derive my existence ? Perhaps from myself, or from my parents, or from some other causes less perfect than God; for anything more perfect, or even equal to God, cannot be thought or imagined.

30. But if I [were independent of every other existence, and] were myself the author of my being, I should doubt of nothing, I should desire nothing, and, in fine, no perfection would be wanting to me; for I should have bestowed upon myself every perfection of which I possess the idea, and I should thus be God. And it must not be imagined that what is now wanting to me is perhaps of more difficult acquisition than that of which I am already possessed; for, on the contrary, it is quite manifest that it was a matter of much higher difficulty that I, a thinking being, should arise from nothing, than it would be for me to acquire the knowledge of many things of which I am ignorant, and which are merely the accidents of a thinking substance; and certainly, if I possessed of myself the greater perfection of which I have now spoken [in other words, if I were the author of my own existence], I would not at least have denied to myself things that may be more easily obtained [as that infinite variety of knowledge of which I am at present destitute]. I could not, indeed, have denied to myself any property which I perceive is contained in the idea of God, because there is none of these that seems to me to be more difficult to make or acquire; and if there were any that should happen to be more difficult to

acquire, they would certainly appear so to me (supposing that I myself were the source of the other things I possess), because I should discover in them a limit to my power.

31. And though I were to suppose that I always was as I now am, I should not, on this ground, escape the force of these reasonings, since it would not follow, even on this supposition, that no author of my existence needed to be sought after. For the whole time of my life may be divided into an infinity of parts, each of which is in no way dependent on any other; and, accordingly, because I was in existence a short time ago, it does not follow that I must now exist, unless in this moment some cause create me anew as it were, that is, conserve me. In truth, it is perfectly clear and evident to all who will attentively consider the nature of duration, that the conservation of a substance, in each moment of its duration, requires the same power and act that would be necessary to create it, supposing it were not yet in existence; so that it is manifestly a dictate of the natural light that conservation and creation differ merely in respect of our mode of thinking [and not in reality].

32. All that is here required, therefore, is that I interrogate myself to discover whether I possess any power by means of which I can bring it about that I, who now am, shall exist a moment afterward: for, since I am merely a thinking thing (or since, at least, the precise question, in the meantime, is only of that part of myself), if such a power resided in me, I should, without doubt, be conscious of it; but I am conscious of no such power, and thereby I manifestly know that I am dependent upon some being different from myself.

33. But perhaps the being upon whom I am dependent is not God, and I have been produced either by my parents, or by some causes less perfect than Deity. This cannot be: for, as I before said, it is perfectly evident that there must at least be as much reality in the cause as in its effect; and accordingly, since I am a thinking thing and possess in myself an idea of God, whatever in the end be the cause of my existence, it must of necessity be admitted that it is likewise a thinking being,

and that it possesses in itself the idea and all the perfections I attribute to Deity. Then it may again be inquired whether this cause owes its origin and existence to itself, or to some other cause. For if it be self-existent, it follows, from what I have before laid down, that this cause is God; for, since it possesses the perfection of self-existence, it must likewise, without doubt, have the power of actually possessing every perfection of which it has the idea—in other words, all the perfections I conceive to belong to God. But if it owe its existence to another cause than itself, we demand again, for a similar reason, whether this second cause exists of itself or through some other, until, from stage to stage, we at length arrive at an ultimate cause, which will be God.

34. And it is quite manifest that in this matter there can be no infinite regress of causes, seeing that the question raised respects not so much the cause which once produced me, as that by which I am at this present moment conserved.

35. Nor can it be supposed that several causes concurred in my production, and that from one I received the idea of one of the perfections I attribute to Deity, and from another the idea of some other, and thus that all those perfections are indeed found somewhere in the universe, but do not all exist together in a single being who is God; for, on the contrary, the unity, the simplicity, or inseparability of all the properties of Deity, is one of the chief perfections I conceive him to possess; and the idea of this unity of all the perfections of Deity could certainly not be put into my mind by any cause from which I did not likewise receive the ideas of all the other perfections; for no power could enable me to embrace them in an inseparable unity, without at the same time giving me the knowledge of what they were [and of their existence in a particular mode].

36. Finally, with regard to my parents [from whom it appears I sprung], although all that I believed respecting them be true, it does not, nevertheless, follow that I am conserved by them, or even that I was produced by them, in so far as I am a thinking being. All that, at the most, they contributed to my

origin was the giving of certain dispositions (modifications) to the matter in which I have hitherto judged that I or my mind, which is what alone I now consider to be myself, is inclosed; and thus there can here be no difficulty with respect to them, and it is absolutely necessary to conclude from this alone that I am, and possess the idea of a being absolutely perfect, that is, of God, that his existence is most clearly demonstrated.

37. There remains only the inquiry as to the way in which I received this idea from God; for I have not drawn it from the senses, nor is it even presented to me unexpectedly, as is usual with the ideas of sensible objects, when these are presented or appear to be presented to the external organs of the senses; it is not even a pure production or fiction of my mind, for it is not in my power to take from or add to it; and consequently there but remains the alternative that it is innate, in the same way as is the idea of myself.

38. And, in truth, it is not to be wondered at that God, at my creation, implanted this idea in me, that it might serve, as it were, for the mark of the workman impressed on his work; and it is not also necessary that the mark should be something different from the work itself; but considering only that God is my creator, it is highly probable that he in some way fashioned me after his own image and likeness, and that I perceive this likeness, in which is contained the idea of God, by the same faculty by which I apprehend myself, in other words, when I make myself the object of reflection, I not only find that I am an incomplete, [imperfect] and dependent being, and one who unceasingly aspires after something better and greater than he is; but, at the same time, I am assured likewise that he upon whom I am dependent possesses in himself all the goods after which I aspire [and the ideas of which I find in my mind], and that not merely indefinitely and potentially, but infinitely and actually, and that he is thus God. And the whole force of the argument of which I have here availed myself to establish the existence of God, consists in this, that I perceive I could not possibly be of such a nature as I am, and yet have in my



mind the idea of a God, if God did not in reality exist—this same God, I say, whose idea is in my mind—that is, a being who possesses all those lofty perfections, of which the mind may have some slight conception, without, however, being able fully to comprehend them, and who is wholly superior to all defect [and has nothing that marks imperfection]: whence it is sufficiently manifest that he cannot be a deceiver, since it is a dictate of the natural light that all fraud and deception spring from some defect.

39. But before I examine this with more attention, and pass on to the consideration of other truths that may be evolved out of it, I think it proper to remain here for some time in the contemplation of God himself—that I may ponder at leisure his marvelous attributes—and behold, admire, and adore the beauty of this light so unspeakably great, as far, at least, as the strength of my mind, which is to some degree dazzled by the sight, will permit. For just as we learn by faith that the supreme felicity of another life consists in the contemplation of the Divine majesty alone, so even now we learn from experience that a like meditation, though incomparably less perfect, is the source of the highest satisfaction of which we are susceptible in this life.

## *Part 12: René*

# *Descartes, Life, Times, and Meditations*

Descartes lived during an intellectually vibrant time. The Scholastics 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 sin-

cere 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*.<sup>115</sup>

To ask “How do we know?” is to ask for reasons that justify our belief in the things we think we know. This is the branch of philosophy called **Epistemology**, which is a fancy word for the study of knowledge (a little meta, I know). The epistemological project of providing systematic justification for the things we take ourselves to know was launched by Descartes and it remains a central endeavor in epistemology to this day. 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.

---

<sup>115</sup>Descartes.

Another area where Descartes has been influential is in the philosophy of mind (Module 3). 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 Elizabeth of Bohemia.

Previously, we explored the philosophy of mind which was launched in the wake of problems for substance dualism and because of the advancements there, Descartes' fruitful failure led to neuroscience, cognitive psychology and information science. 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

The final big issue that Descartes brought enduring attention to is the problem of free will (Module 4). 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 nettle 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 our bodily movements.

# *Part 13: An Overview of Descartes' Meditations on First Philosophy*

**Epistemology** (from the Greek word (episteme) meaning 'knowledge') is a branch of philosophy which deals with knowledge and belief. Some of the questions which are found there are:

1. What does it take to know something?
2. What is belief?
3. What if faith? (This is a more recent question)
4. What are the different kinds of knowledge? Are some better than others?
5. What is wisdom?

6. What separates knowledge from opinion?
7. And there are many, many others.

Despite what others may tell you, and I have debated this on many occasions, Descartes' Meditations concern epistemology, not metaphysics. So, 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 has been 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.

The meditations are broken up into 6 parts and each serves a different purpose. I encourage you to read all 6, but the first 2 are enough for this module.

## The First Meditation

Below is the first small section of this meditation, written in more plain language. The point of these meditations is to break down everything, doubt everything, until we can find something which can't be doubted and use that as the basis for building up a 100% guaranteed true set of beliefs/knowledge. The first meditation concerns what can be doubted.

Several years ago, I came to the opinion that many of the things which I believed were true, were, in fact, false; and when that happened, I thought that I should do the job of trying to rid myself of these false opinions in order to get an indubitable structure for knowledge (science). But I waited until I was older and wiser so that I could have very little doubt in my abilities. On this account, I have delayed so long that it would be wrong to keep thinking on it and I should just put pen to paper. Today, then, since I don't have anything which I need to do because I am retired, I will at length apply myself earnestly and freely to the general doubting of all my former opinions.

For the second section of this meditation, Descartes starts us off by explaining his method for getting at that which cannot be doubted. Rather than dealing with each belief individually, he chooses to deal with classes, or categories, of beliefs (such as ones which all came from the senses, or were told to him by some person). If he finds reason to doubt some of the beliefs in a class, then he can doubt the entire class. The type of reason to doubt is also important. The reason must be one which is at the foundation of those beliefs. If you take out the foundation, the rest collapses. Here is an example of this general idea in practice:

Take the beliefs that you got because someone told them to you. But, as we all know, people do sometimes lie or are mistaken about what they are telling others. This means that there is reason to doubt all of the things which you were taught in school, or by your folks, or by other people. Since there is reason to doubt those things, they can't be the foundation of indubitable knowledge.

Given the times Descartes was living in, with the rise of empirical science (the idea that knowledge is primarily gained from



the senses, through experience and experiments), Descartes wanted to show that the fundamental basis for this knowledge was actually grounded in knowledge you can gain without appealing to your senses, called *a priori* knowledge. So, he turns to whether one can doubt their senses. To get this off the ground, Descartes needs a way of getting rid of knowledge through your senses, as in, knowledge via experience. This means he needs a systematic way of making a worst case scenario (think Murphy's Law, if it can go wrong it will), making it so that the foundation of his knowledge is internal, not external. According to Descartes, you can doubt your senses, meaning that they can't be at the foundation of an indubitable knowledge structure. He gives us two ways, with the second stronger than the first.

1. What if he is having a really realistic dream? (this is in sections 6-8)
2. What if he is under the influence of some evil demon? (this is in sections 9-12)

The third section of this meditation gives us a basis for making the claim that our senses cannot be the foundational basis for knowledge. Put in more plain language, he says:

Everything that I have held with certainty, I have gotten through my senses. But, I know that my senses have sometime misled me; as a general rule, you should not trust something (100%) if it has lied to you.

Every generation or so has a movie or show which draws inspiration from Descartes' Meditations or Plato's Cave (they both use similar thought experiments, but draw different, though related, conclusions. One great example is the first Matrix movie. So much inspiration was drawn that we have this little quote from Morpheus to Neo:

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

The point here is that we could be deceiving ourselves, what if we are having a really, super, realistic dream? It would seem that, then, we couldn't trust our senses, rather what we experience would be an illusion.

### **The dream case (this is section 6)**

At this point, Descartes gives the first thought experiment, or possible scenario, where his senses could be misleading, allowing him to discard them as the foundation for knowledge. Put more plainly, this section goes like this:

Hey, you know, I am a man and, you know, I need to sleep. When I sleep, I tend to dream and be in those dreams. Sometimes those dreams are really crazy, but other times they make a lot of sense. Sometimes, I cannot tell the difference between my dreams and reality. How can I be sure that I am not, right now, dreaming something really realistic?

This case, however, is not strong enough for Descartes' purposes. If he were in a very realistic dream, then Descartes could still gain some knowledge from his senses, but this knowledge would be fairly abstract. Here is the plain language version of his reasoning:

So, let's suppose that I am having this really realistic dream. This would mean that everything I am seeing now are nothing more than illusions from the unconscious mind. But we have to admit that the things in my dreams are based on things in the real world, no matter how mangled and jumbled they may be. Because of this, we can say that although the objects in dreams may be doubted, we must claim that there are things which they are based on. So, you can still not doubt your experienced whole-hog, like I'm trying to do.

As you can see, if he were in a very realistic dream, then he would still have some knowledge about the external world. This is because dreams are based on experiences in the waking world. One can still have knowledge of abstract things like shape, color, motion, and the like. In order to completely discount knowledge gained from the senses, he needs a stronger possible case.

## God and Deception

He believes that there is an all good God, but how could this fit with being mistaken in this way? If allowing a person to be mistaken did not jive with God, He would not allow it. But He does, we are mistaken all the time, so allowing a person to be mistaken must be just fine by God.

## The Evil Demon Case (this is section 12)

Here we have the Evil Demon, which is the thought experiment strong enough to discredit all knowledge gained from the senses. This thought experiment is the basis for many books, movies, and shows.<sup>116</sup> In plain language, his introduction of this case goes like this:

---

<sup>116</sup>need to list examples

I will suppose that some Evil Demon has employed all of their power to deceive me. All of my senses and the things which I learned from them are nothing more than either illusions or are the product of the Demon's deceptions. I will consider myself as without a body or anything external to the mind, as they are, too, illusions by the Demon. But this this is a really hard job, and difficult to maintain; so I tend to fall back into thinking that the external world is indubitable when I leave my armchair. This tendency does not mean that they are, in fact, indubitable, but rather, that we have a built in bias.

## Descartes' Second Meditation

The last meditation left me with no knowledge at all, everything could be doubted, and I have a hard time forgetting this. Also, I don't see a way out of the situation, it is like I have been plunged in deep water and can't find my way up or down. Despite this, I will forge on, treating everything with the slightest doubt of its truth as if it were false. I will continue to do this until I have found something which I know for certain or, at the very least, that I know for certain that nothing else is know for certain. For Archimedes to move the world from one place to another only demanded a firm and immovable point. In the same way, if I can find one thing which is indubitable, then I can have great hopes for this project.

Right now, I suppose that everything I see are illusions or are not actual. I can't trust my memory, so I must assume that the things I remember never existed. My senses are also dubitable, so I must assume that I don't have them. Because I learned about things like body, figure, extension and motion from my senses, I must assume that they are not actual as well. Since I can't have any of those things, what can I have which is true? Maybe only that nothing is certain.

How do I know that there isn't something different than those things which I can't doubt? Didn't I suppose that there was some kind of Demon who could cause these thoughts to come to my mind? But why do I even need to bring up this Demon, when I would deceive myself. Even if there is some demon, I am still something, I still exist. I have supposed that I don't have a body, but then I must ask what comes from that? Am I so dependent on the body and the senses that without them, I don't exist? But I had been persuaded that there was nothing in the world. When I was persuaded, was I also persuaded that I don't exist? Not at all, I certainly exist, since I was persuaded. At the same time, I supposed that there was some great demon deceiving me. This further shows that I exist, since I am being deceived. No matter how much this demon deceives me, he can't make it that I don't exist, so long as I am conscious, I am something. There is at least one thing which I cannot doubt, I am, I exist.

Even though I know that I exist, I don't know what I am. The next step is one I need to take care in making, because if I say that I am the wrong sort of thing, then I will have strayed from the truth. This means that I need to think about what I thought I was and cast those aside. I will apply the same methods as before in order to get at what is undoubtable.

But what can I say I am if I suppose that there is this demon? Can I say that I still have all of the things which belong to the body? After looking into it, I can say that done really belong to me. Let's move on, then, to the attributes of the soul. The first mentioned were the powers of nutrition and walking; but, if I don't have a body, then I can't be able to either walk or be nourished. Second, we had perception. In the same way as before, perception requires that I have a body, besides, I have often thought my self having perceptions in dreams. The third and last attribute is thinking; and this is what really belongs to me. This alone is inseparable from me. I am—I exist: this is certain; but, when do I exist? I exist whenever I am thinking. If I stop thinking, it would seem that I cease to exist. I am therefore only a thinking thing, that is, a mind. I am, however, a real thing, and really existing; but what kind of thing? The answer was a thinking thing.

### Commentary:

Here, Descartes has his Evil Demon case, think of it like the Matrix without the mad glitches. This puts him in a state where he can doubt almost everything. Descartes ends up finding one thing which he can't doubt.

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 too.

## Descartes' Third Meditation

At this point, Descartes only has two justified beliefs. One is about his thoughts (the content of his mind) and the other is about his own existence (I think therefore I am). At the moment, he doesn't have any knowledge, or at least justified beliefs about anything outside of his mind. For example, he can't say "I know I have hands" or "I know there's a piece of paper in front of me."

Also, Descartes doesn't have any information about truths we know through reason alone, like  $2+2=4$  or other things like that. Those just don't have the justification yet. To get some basis to justify that stuff, Descartes sets out to prove that God exists and is not like his evil demon. Once the evil deceiver hypothesis is taken out because of God, facts we get from reason and perhaps those from our senses could get the justification to be knowable. However, not just any argument for the existence of God and that God's all-good will do the trick (we will be looking into a lot of different arguments for the existence of God in Module 5). The trick for Descartes' project of a rational reconstruction of knowledge is to prove the existence of a good God by reasoning using only those beliefs that he has identified as indubitable and foundational.

Descartes argument for the existence of a good God goes roughly as follows:

1. I find in my mind the idea of a perfect being.
2. The cause of my idea of a perfect being must have at least as much perfection and reality as I find in the idea.
3. I am not that perfect.
4. Nothing other than a good and perfect God could be the cause of my idea of a perfect being.
5. So, a good and perfect God must exist.



This argument is a really simple version of the argument which Descartes gives (it's deep and complex) but the core features are there and can show the flaw in his reasoning (all arguments for the existence of God have at least one flaw (most of them have the same flaw, as we will see). Let's grant that this argument is valid and move over to look at it's soundness. Keep in mind that for Descartes to do the job he is trying to, all of his premises in this argument need to be indubitable and foundational. The first premise is a belief about his own thoughts, so we will give it to him. Though it is not as clear, premise three might arguably count as a foundational belief about the contents of Descartes' own mind. An evil deceiver, being evil, would lack perfection found in Descartes idea of a perfect being. So as powerful as such a being could be, the cause of Descartes idea of a perfect being must be more perfect than any evil deceiver. Perhaps any being so perfect would have to be a good God.

But the issue for Descartes is in the second premise. What reason do we have for thinking that the cause of something must have at least as much perfection as its effect?

The idea that there are degrees of perfection and the notion that the less perfect can only be explained in terms of the more perfect traces it's way back to Plato and his theory of Forms. Like many ideas of this type, it will strike many of us as implausible or even incomprehensible. What, exactly, is perfection supposed to mean here? And even once we've spelled this out, why think causes must be more perfect?

It doesn't seem all that rare for less perfect things to make more perfect things (if you are a parent, it might seem clear that your kid is more perfect (better) than you are, or the ugly depressed artist making a beautiful painting).

In any case, whether the second premise can be explained and defended at all, the fatal flaw for Descartes' project is that it is not foundational. It is not an indubitable belief about the contents of Descartes' own mind, but rather a substantive belief about how things are beyond the bounds of Descartes' own mind. So Descartes' attempt to provide a rational justification

for a substantive body of knowledge leaves us with an enduring skeptical problem. All we have immediate intellectual access to is the contents of our own minds. How can we ever have knowledge of anything beyond the contents of our own mind based on this? This is the problem of Cartesian skepticism.

Having diagnosed the fatal flaw in Descartes' project, we should briefly consider how his rational reconstruction of knowledge was to go from there. Given knowledge of God's existence and good nature, we would appeal to this to assure the reliability of knowledge had through reason and later also through the senses. God being the most perfect and good being would rule out the possibility of interference by an evil deceiver. We might still make mistakes in reasoning or be misinformed by the senses. But this would be due to our failure to use these faculties correctly. A good God, however, would not equip us with faculties that could not be trusted to justify our beliefs if used properly. This is a very cursory summary of the later stages of Descartes' attempted rational reconstruction of knowledge in his *Meditations*. But it will suffice for our purposes.

# *Part 14: Knowledge and Justification*

In the previous section of this module, we were discussing René Descartes and his quest for certainty and knowledge. He did this by employing a form of methodological skepticism, doubt everything which can be doubted. In philosophy, **knowledge**, as we are going to use it, is justified true belief. While some claim that it is possible to know something without believing it and others claim that it is possible to know something false, these stances are pretty easily pushed aside. There are several ways in which one could be justified, for various things. For example, being entitled to your medical records makes you justified in requesting them. However, the sort of justification seen when it comes to knowledge is different. Epistemic justification is the kind found when it comes to knowledge. This is background reasoning which supports the belief. A belief is epistemically justified if we have sufficiently good reason to think that the belief is true. Much of the debate in Epistemology centers around the standards for justification. If you believe something and that thing happens to be true, what else is needed for you to know it?

Some of you might think that you just need to be right to know it. However, if you are like most of us, there have been

times when you were correct purely by chance. For example, recently, my little brother came to me with a question about Dragon Ball Z and the time travel in it. As an expert on theories of time travel, I thought this was great. He came up with the idea that there had to be four different timelines in the show, given what he had seen thus far. Explaining that there had to be two different versions of the same character and so forth. As it turns out, he was correct, there did need to be two different versions of that character and there needed to be 4 timelines, but for completely different reasons than the ones he laid out. He later said that it didn't feel like he was correct, something was missing. I quickly recognized what was missing: adequate **justification**. The structure of his reasoning, though it lead to the correct answer, wasn't accurate.

Examples like this can be found all over the place, where you were correct, but for totally unexpected reasons. This is why good critical thinkers look deeply at the structure of their justification, how their beliefs relate to each other, which also calls back to one of Socrates' core questions; what is your evidence?

Here is a, relatively, classic example of justification going astray. There have been various thought experiments which follow this line of thought, but I will go with the one which is earliest, as far as I am aware, from Alexius Meinong:

Imagine that you live near a park and most days at noon, a violinist plays. You can hear the beautiful music from your dinning room and, it seems right to say that if you hear the music, then you know that the violinist is playing. However, one day, because you, unknowingly, drank some bad tea, at noon, you go temporarily deaf and have an auditory hallucination of a violin playing. While it is true that the violinist is out there and you hear the music, does it follow that you know the violinist is out there?

For cases like these, most people would say that, though you

were right, you don't know that the violinist is playing. The reason, as before, is that there was something wrong with the justification. It should be noted that 100% certainty, no chance of being wrong, is not necessarily required for justification. For example, in a real world case, you can know that your sink isn't leaking without being in the room with the sink. There is a chance you are wrong, given your evidence. Justification can come in degrees and the more justified a belief is, then the more likely it is to be accurate/true. In the violinist example I gave, experience and prior evidence certainly gives ample justification for knowledge in most cases, but you would have had the same evidence if the violinist had been sick and unable to play. This is why good critical thinkers question, doubt, or even reject beliefs if they lack adequate justification.

Justification for knowledge is always how your beliefs relate to each other. You can think of this as a building or a puzzle (these ways of picturing the structure lead to two different theories of justification). After doing just a little introspection, you will find that your beliefs relate to each other in various ways, and some make others more likely to be accurate. For example, take this case:

My belief that the weather-forecast in Washington is unreliable is based on my belief that they have claimed many times that it would be sunny and on my belief that the weather on many of those occasions was, in fact, rainy. My belief about their claims is based on my belief that I remember their statements correctly. My belief about the weather on those occasions is based on my belief that my memory is correct.

For this, we see that my beliefs form a pyramid, or tower, like structure, which leads us right into the first theory of justification, Foundationalism, which is what Descartes used in the *Meditations*.

## Foundationalism

Much of the debate in Epistemology, as I mentioned, centers around justification. There are three dominate theories of justification which we will cover. Doing a little introspection should make it clear that some beliefs are built upon others. One belief is justified by another belief. For example, (and this is a strange system) my belief is that the Earth is actually a dodecahedron is supported (justified) by my beliefs that the middle of two extremes is most likely accurate and when it comes to the shape of the Earth, one extreme says that it is round, and the other says it is flat. Though the beliefs in question are only partially accurate, the middle of two extremes is not always the most accurate, the justification flows in that way. The debate about justification is how the beliefs either need to relate to each other or the sort of cement used to link them together.

The most basic and common stance in Epistemology is **Foundationalism**. This is the stance that our beliefs, in order to be justified, need to be ultimately justified on the basis of certain, core, foundational beliefs. For example, my belief that the weather-forecast in Washington is unreliable is based on my belief that they have claimed many times that it would be sunny and on my belief that the weather on many of those occasions was, in fact, rainy. My belief about their claims is based on my belief that I remember their statements correctly. My belief about the weather on those occasions is based on my belief that my memory is correct. At the bottom of this structure are the foundational beliefs, which support all of the others. Your beliefs, if foundationalism is correct, can't form a circle of justification. For example, suppose that my belief that Mt. Rainer is an active volcano is supported by my belief that what my teacher in middle school told me was accurate. That belief is supported by the belief that she got accurate knowledge from her schooling and that is further supported by the belief that it's accurate that Mt. Rainer is an active volcano. If we follow the train back, we notice that my belief that Mt.

Rainier is an active volcano is, at least indirectly, supported by itself. That should seem strange and the foundationalist wants to exclude cases like that from leading to knowledge.

Foundational beliefs are your rock bottom, these aren't supported or justified by anything else. If your beliefs are structured in this way and you want to get at knowledge, certainty, then you need to look very closely at the foundational beliefs. These beliefs need to be self-evident, they are not supported by other beliefs, they need to be strong enough to carry all the weight, all by themselves. If you have any experience with construction, or Minecraft, this analogy will work nicely for you. The foundationalist thinks of the intellectual structure of knowledge, how the beliefs relate to each other, as a tall building. Buildings can take many different shapes, from tall towers to pyramids, to strange Brutalist architecture, with a small foundation and an expanding roof. But, at the end of the day, the foundation, the concrete slab it's built on, needs to be tough enough to take the weight.

Finding your foundational beliefs might require a little help, or a highly questioning mind. Little kids will often continually ask 'why'. It could start with something fairly simple like 'mommy, why is the sky blue?' and then you will answer with something accurate to the best of your knowledge, and then they will ask 'why', you will reply, and they will ask 'why' again, and again, and again. Eventually, you will get to a point where you just need to throw your hands up and say 'that's just the way it is.' For example, last time I had this happen, I just stopped at 'because somethings exist and other things don't.' This final belief is your foundation. If you keep on digging, there's nothing left. Is this strong enough, on its own, to hold up the weight? In my weather forecast example, the foundations could be that my senses and memory are accurate (for those cases). Is that strong enough, all on its own?

Foundational beliefs tend to be very basic logical or experiential truths. For example, something like 'my senses are accurate' or 'if P implies Q and P is true, then Q is true', or that

there can't be contradictions. These all can be used to build up a structure of beliefs, which, though possible to be inaccurate, still count as justification.

## Infinite Regress Theory

The **Infinite Regress Theory** of justification points the finger at the foundational beliefs. For a belief to count as knowledge, according to the foundationalist, need to be built up from basic, foundations, which are self-evident. Those beliefs don't require any further justification. This theory claims that all beliefs need to be justified. It agrees with foundationalism that beliefs can't directly or indirectly support themselves, but it claims that foundational beliefs can't lead to knowledge. For a belief to count as knowledge, according to this theory, it must be justified by some other beliefs, which in turn need to be further justified by others, and those need to be justified by still others, and so on. For you to know something, there would need to be an infinite string of justifying beliefs, never making a circle, which support it. Such an account of justification makes knowledge, frankly, impossible.

Some of you might have heard the phrase 'it's turtles all the way down', which has a humorous story behind it. This story has changed and mutated quite a bit with time, but it gets the general idea:



Long ago, a philosopher was giving a lecture to the public. In it, they mention the structure of the solar system and the planets. Afterwards, an older lady comes up to the philosopher and says "your theory that the sun is at the center and the Earth is a ball is very convincing, but I have a better one." Amused, the philosopher asks about it and she replies "we live on the crust of the Earth which rests on a turtle's back." Instead of demolishing her stance with evidence, the philosopher chose to follow Socrates and ask follow-up questions, "if that is correct, ma'am, then what does the turtle stand on?" Equally amused, the lady replied "the first turtle stands on the back of a second turtle." "But," replied the philosopher, "what does the second turtle stand on?" She then said "why, the back of yet another, bigger turtle, it's no use, Mister, it's turtles all the way down."

Infinite Regress Theory follows a very similar line of thought. Rather than their being foundational beliefs, in order to have knowledge, there must be beliefs all the way down. However, having an infinite chain of knowledge doesn't, on its own, give us knowledge. For example, suppose that there's at least 0 soda cans on my desk. Here is a chain of beliefs; if there's at least 2 soda cans on my desk, then there's at least 1; if there are 3 soda cans on my desk then there are at least 2; if there are 4... Continuing for all numbers. None of these beliefs are really justified, the chain will never reach back to reality and justify my claim that there's no soda cans on my desk. This also means that, if this theory is correct, knowledge is impossible and Global Skepticism is the way to go.

## Coherentism

Building off of the previous two, Foundationalism claims that there are bedrock beliefs which don't require further justifica-

tion, knowledge is justified on the basis of other beliefs (aside from the foundational ones), and beliefs can't directly or indirectly justify themselves to count as knowledge. Infinite Regress Theory claims that there are no foundational beliefs, knowledge must be justified on the basis of further more basic beliefs, and beliefs can't directly or indirectly justify themselves. **Coherentism** is like Infinite Regress Theory in that it claims that there are no foundational beliefs in the structure of knowledge, but it rejects that beliefs can't indirectly justify themselves. Beliefs are justified in how they cohere or fit together. It is a sort of back and forth process. Justification can come in two different directions. For Foundationalism and Infinite Regress, it only works in one direction, basic on up. All three agree that seeing many photos of the Earth and, even, going up in a space shuttle would justify the belief that the Earth is round. In such a case, many individual observations build up evidence for a more grand generalized claim. The particular observations give us a basis for the principles. However, sometimes the generalized beliefs can justify our particular evidence. The generalized claims stick around by categorizing or making sense of our observations.

Sometimes, we have cases where some evidence, an observation, just can't fit in the generalization. Sometimes, the evidence disproves our generalization. For example, suppose that after years of bird watching in states aside from Washington and seeing thousands of ravens, I generalize to the belief that all ravens are black. But, one day, while bird-watching in Washington, I spot two white ravens (yes, there is a beach in Washington with a family of white ravens). This will clearly conflict with my core belief that all ravens are black, but that's OK. After confirming that those are in fact ravens and doing my due diligence, I amend my belief to something like "most ravens are black."

But that is still fine with the foundationalist, they would say that the process is still bottom up. The coherentist will make a further claim that sometimes our general theories shape our

observations. This can be because of the biases which we have covered in this class or it could be because of other reasons. Coherentism claims that our observations can be 'theory laden', or influenced by our general theories. This is not a bottom-up process, like the foundationalist claims, rather it's a 'top-down'. The coherentist would prefer that we think of justification not as a tower, or building, but as a flat spider-web or a mat. For example, a person, Mary, strongly believes in miracles. When she sees, for example, a statue of Jesus with water dripping from the cheeks, Mary will instantly color her experience of the 'weeping statue' as a miracle. Similarly, Norm, who strongly holds that miracles don't happen, will see such an event and reject it almost instantly as his eyes playing tricks on him or will jump to some explanation in order to explain the event. This is the kind of back and forth which Coherentism entails. People who strongly believe in Sasquatch are likely to see Sasquatch because their beliefs color their experience and their experience shapes their beliefs.

It is worth noting that without an infinite number of beliefs, which no person can have, and without foundational beliefs, it's not possible to have a system of beliefs fit the coherentist model without there being some kind of circle, at least one. In Mary's case, her belief that the weeping statue is a miracle justifies her stance that miracles happen and her stance that miracles happen justifies her stance that it was a miracle.

## MODULE VII

### *Is Morality Culturally Relative?*

# *The Challenge of Cultural Relativism by James Rachels*

117

“Morality differs in every society, and is a convenient term for socially approved habits.” Ruth Benedict, *Patterns of Culture* (1934)

## **2.1 How Different Cultures Have Different Moral Codes**

Darius, a king of ancient Persia, was intrigued by the variety of cultures he encountered in his travels. He had found, for example, that the Callatians (a tribe of Indians) customarily ate the bodies of their dead fathers. The Greeks, of course,

---

<sup>117</sup>James Rachels, “*The Challenge of Cultural Relativism*,” *Exploring Philosophy: An Introductory Anthology*, edited by Steven M. Cahn (Oxford UP, 1907).

did not do that—the Greeks practiced cremation and regarded the funeral pyre as the natural and fitting way to dispose of the dead. Darius thought that a sophisticated understanding of the world must include an appreciation of such differences between cultures. One day, to teach this lesson, he summoned some Greeks who happened to be present at his court and asked them what they would take to eat the bodies of their dead fathers. They were shocked, as Darius knew they would be, and replied that no amount of money could persuade them to do such a thing. Then Darius called in some Callatians, and while the Greeks listened asked them what they would take to burn their dead fathers' bodies. The Callatians were horrified and told Darius not even to mention such a dreadful thing.

This story, recounted by Herodotus in his *History* illustrates a recurring theme in the literature of social science: Different cultures have different moral codes. What is thought right within one group may be utterly abhorrent to the members of another group, and vice versa. Should we eat the bodies of the dead or burn them? If you were a Greek, one answer would seem obviously correct; but if you were a Callatian, the opposite would seem equally certain.

It is easy to give additional examples of the same kind. Consider the Eskimos. They are a remote and inaccessible people. Numbering only about 25,000, they live in small, isolated settlements scattered mostly along the northern fringes of North America and Greenland. Until the beginning of the 20th century, the outside world knew little about them. Then explorers began to bring back strange tales.

Eskimos customs turned out to be very different from our own. The men often had more than one wife, and they would share their wives with guests, lending them for the night as a sign of hospitality. Moreover, within a community, a dominant male might demand and get regular sexual access to other men's wives. The women, however, were free to break these arrangements simply by leaving their husbands and taking up with new partners—free, that is, so long as their former hus-

bands chose not to make trouble. All in all, the Eskimo practice was a volatile scheme that bore little resemblance to what we call marriage.

But it was not only their marriage and sexual practices that were different. The Eskimos also seemed to have less regard for human life. Infanticide, for example, was common. Knud Rasmussen, one of the most famous early explorers, reported that he met one woman who had borne 20 children but had killed 10 of them at birth. Female babies, he found, were especially liable to be destroyed, and this was permitted simply at the parents' discretion, with no social stigma attached to it. Old people also, when they became too feeble to contribute to the family, were left out in the snow to die. So there seemed to be, in this society, remarkably little respect for life.

To the general public, these were disturbing revelations. Our own way of living seems so natural and right that for many of us it is hard to conceive of others living so differently. And when we do hear of such things, we tend immediately to categorize those other peoples as "backward" or "primitive." But to anthropologists and sociologists, there was nothing particularly surprising about the Eskimos. Since the time of Herodotus, enlightened observers have been accustomed to the idea that conceptions of right and wrong differ from culture to culture. If we assume that our ideas of right and wrong will be shared by all peoples as all times, we are merely naive.

## 2.2 Cultural Relativism

To many thinkers, this observation—"Different cultures have different moral codes"—has seemed to be the key to understanding morality. The idea of universal truth in ethics, they say, is a myth. The customs of different societies are all that exist. These customs cannot be said to be "correct" or "incorrect," for that implies we have an independent standard of right and wrong by which they may be judged. But there is no such

independent standard; every standard is culture-bound. The great pioneering sociologist William Graham Sumner, writing in 1906, put the point like this:

The "right" way is the way which the ancestors used and which has been handed down. The tradition is its own warrant. It is not held subject to verification by experience. The notion of right is in the folkways. It is not outside of them, of independent origin, and brought to test them. In the folkways, whatever is, is right. This is because they are traditional, and therefore contain in themselves the authority of the ancestral ghosts. When we come to the folkways we are at the end of our analysis

This line of thought has probably persuaded more people to be skeptical about ethics than any other single thing. Cultural Relativism, as it has been called, challenges our ordinary belief in the objectivity and universality of moral truth. It says, in effect, that there is not such thing as universal truth in ethics; there are only the various cultural codes, and nothing more. Moreover, our own code has no special status; it is merely one among many.

As we shall see, this basic idea is really a compound of several different thoughts. It is important to separate the various elements of the theory because, on analysis, some parts turn out to be correct, while others seem to be mistaken. As a beginning, we may distinguish the following claims, all of which have been made by cultural relativists:

- 1 Different societies have different moral codes.
- 2 There is no objective standard that can be used to judge one societal code better than another.
- 3 The moral code of our own society has no special status; it is merely one among many.



- 4 There is no "universal truth" in ethics; that is, there are no moral truths that hold for all peoples at all times.
- 5 The moral code of a society determines what is right within that society; that is, if the moral code of a society says that a certain action is right, then that action is right, at least within that society.
- 6 It is mere arrogance for us to try to judge the conduct of other peoples. We should adopt an attitude of tolerance toward the practices of other cultures.

Although it may seem that these six propositions go naturally together, they are independent of one another, in the sense that some of them might be false even if others are true. In what follows, we will try to identify what is correct in Cultural Relativism, but we will also be concerned to expose what is mistaken about it.

## 2.3 The Cultural Differences Argument

Cultural Relativism is a theory about the nature of morality. At first blush it seems quite plausible. However, like all such theories, it may be evaluated by subjecting it to rational analysis; and when we analyze Cultural Relativism we find that it is not so plausible as it first appears to be.

The first thing we need to notice is that at the heart of Cultural Relativism there is a certain form of argument. The strategy used by cultural relativists is to argue from facts about the differences between cultural outlooks to a conclusion about the status of morality. Thus we are invited to accept this reasoning:

- 1 The Greeks believed it was wrong to eat the dead, whereas the Callatians believed it was right to eat the dead.

- 2 Therefore, eating the dead is neither objectively right nor objectively wrong. It is merely a matter of opinion, which varies from culture to culture.

Or, alternatively:

- 1 The Eskimos see nothing wrong with infanticide, whereas Americans believe infanticide is immoral.
- 2 Therefore, infanticide is neither objectively right nor objectively wrong. It is merely a matter of opinion, which varies from culture to culture.

Clearly, these arguments are variations of one fundamental idea. They are both special cases of a more general argument, which says:

- 1 Different cultures have different moral codes.
- 2 Therefore, there is no objective "truth" in morality. Right and wrong are only matters of opinion, and opinions vary from culture to culture.

We may call this the Cultural Differences Argument. To many people, it is persuasive. But from a logical point of view, is it sound?

It is not sound. The trouble is that the conclusion does not follow from the premise—that is, even if the premise is true, the conclusion still might be false. The premise concerns what people believe. In some societies, people believe one thing; in other societies, people believe differently. The conclusion, however, concerns what really is the case. The trouble is that this sort of conclusion does not follow logically from this sort of premise.

Consider again the example of the Greeks and Callatians. The Greeks believed it was wrong to eat the dead; the Callatians believed it was right. Does it follow, from the mere fact that they disagreed, that there is no objective truth in the matter?

No, it does not follow; for it could be that the practice was objectively right (or wrong) and that one or the other of them was simply mistaken.

To make the point clearer, consider a different matter. In some societies, people believe the earth is flat. In other societies, such as our own, people believe the earth is (roughly) spherical. Does it follow, from the mere fact that people disagree, that there is no "objective truth" in geography? Of course not; we would never draw such a conclusion because we realize that, in their beliefs about the world, the members of some societies might simply be wrong. There is no reason to think that if the world is round everyone must know it. Similarly, there is no reason to think that if there is moral truth everyone must know it. The fundamental mistake in the Cultural Differences Argument is that it attempts to derive a substantive conclusion about a subject from the mere fact that people disagree about it.

This is a simple point of logic, and it is important not to misunderstand it. We are not saying (not yet, anyway) that the conclusion of the argument is false. It is still an open question whether the conclusion is true or false. The logical point is just that the conclusion does not follow from the premise. This is important, because in order to determine whether the conclusion is true, we need arguments in its support. Cultural Relativism proposes this argument, but unfortunately the argument turns out to be fallacious. So it proves nothing.

## 2.4 The Consequences of Taking Cultural Relativism Seriously

Even if the Cultural Differences Argument is invalid, Cultural Relativism might still be true. What would it be like if it were true?

In the passage quoted above, William Graham Sumner summarizes the essence of Cultural Relativism. He says that there

is no measure of right and wrong other than the standards of one's society: "The notion of right is in the folkways. It is not outside of them, of independent origin, and brought to test them. In the folkways, whatever is, is right." Suppose we took this seriously. What would be some of the consequences?

### **1. We could no longer say that the customs of other societies are morally inferior to our own.**

This, of course, is one of the main points stressed by Cultural Relativism. We would have to stop condemning other societies merely because they are "different." So long as we concentrate on certain examples, such as the funerary practices of the Greeks and Callatians, this may seem to be a sophisticated, enlightened attitude.

However, we would also be stopped from criticizing other, less benign practices. Suppose a society waged war on its neighbors for the purpose of taking slaves. Or suppose a society was violently anti-Semitic and its leaders set out to destroy the Jews. Cultural Relativism would preclude us from saying that either of these practices was wrong. We would not even be able to say that a society tolerant of Jews is better than the anti-Semitic society, for that would imply some sort of transcultural standard of comparison. The failure to condemn these practices does not seem enlightened; on the contrary, slavery and anti-Semitism seem wrong wherever they occur. Nevertheless, if we took Cultural Relativism seriously, we would have to regard these social practices as also immune from criticism.

### **2. We could decide whether actions are right or wrong just by consulting the standards of our society.**

Cultural Relativism suggests a simple test for determining what is right and what is wrong: All one need do is ask whether the action is in accordance with the code of one's society. Sup-

pose in 1975, a resident of South Africa was wondering whether his country's policy of apartheid—a rigidly racist system—was morally correct. All he has to do is ask whether this policy conformed to his society's moral code. If it did, there would have been nothing to worry about, at least from a moral point of view.

This implication of Cultural Relativism is disturbing because few of us think that our society's code is perfect; we can think of ways it might be improved. Yet Cultural Relativism would not only forbid us from criticizing the codes of other societies; it would stop us from criticizing our own. After all, if right and wrong are relative to culture, this must be true for our own culture just as much as for other cultures.

### **3. The idea of moral progress is called into doubt.**

Usually, we think that at least some social changes are for the better. (Although, of course, other changes may be for the worse.) Throughout most of Western history the place of women in society was narrowly circumscribed. They could not own property; they could not vote or hold political office; and generally they were under the almost absolute control of their husbands. Recently much of this has changed, and most people think of it as progress.

If Cultural Relativism is correct, can we legitimately think of this as progress? Progress means replacing a way of doing things with a better way. But by what standard do we judge the new ways as better? If the old ways were in accordance with the social standards of their time, then Cultural Relativism would say it is a mistake to judge them by the standards of a different time. Eighteenth-century society was, in effect, a different society from the one we have now. To say that we have made progress implies a judgment that present-day society is better, and that is just the sort of transcultural judgment that, according to Cultural Relativism, is impermissible.

Our idea of social reform will also have to be reconsidered. Reformers such as Martin Luther King, Jr., have sought to change their societies for the better. Within the constraints imposed by Cultural Relativism, there is one way this might be done. If a society is not living up to its own ideals, the reformer may be regarded as acting for the best: The ideals of the society are the standard by which we judge his or her proposals as worthwhile. But the "reformer" may not challenge the ideals themselves, for those ideals are by definition correct. According to Cultural Relativism, then, the idea of social reform makes sense only in this limited way.

These three consequences of Cultural Relativism have led many thinkers to reject it as implausible on its face. It does make sense, they say, to condemn some practices, such as slavery and anti-Semitism, wherever they occur. It makes sense to think that our own society has made some moral progress, while admitting that it is still imperfect and in need of reform. Because Cultural Relativism says that these judgments make no sense, the argument goes, it cannot be right.

## 2.5 Why There Is Less Disagreement Than It Seems

The original impetus for Cultural Relativism comes from the observation that cultures differ dramatically in their views of right and wrong. But just how much do they differ? It is true that there are differences. However, it is easy to overestimate the extent of those differences. Often, when we examine what seems to be a dramatic difference, we find that the cultures do not differ nearly as much as it appears.

Consider a culture in which people believe it is wrong to eat cows. This may even be a poor culture, in which there is not enough food; still, the cows are not to be touched. Such a society would appear to have values very different from our own. But does it? We have not yet asked why these people will not

eat cows. Suppose it is because they believe that after death the souls of humans inhabit the bodies of animals, especially cows, so that a cow may be someone's grandmother. Now do we want to say that their values are different from ours? No; the difference lies elsewhere. The difference is in our belief systems, not in our values. We agree that we shouldn't eat Grandma; we simply disagree about whether the cow is (or could be) Grandma.

The point is that many factors work together to produce the customs of a society. The society's values are only one of them. Other matters, such as the religions and factual beliefs held by its members, and the physical circumstances in which they must live, are also important. We cannot conclude, then, merely because customs differ, that there is a disagreement about values. The difference in customs may be attributable to some other aspects of social life. Thus there may be less disagreement about values than there appears to be.

Consider again the Eskimos, who often kill perfectly normal infants, especially girls. We do not approve of such things; a parent who killed a baby in our society would be locked up. Thus there appears to be a great difference in the values of our two cultures. But suppose we ask why the Eskimos do this. The explanation is not that they have less affection for their children or less respect for human life. An Eskimo family will always protect its babies if conditions permit. But they live in a harsh environment, where food is in short supply. A fundamental postulate of Eskimos thought is: "Life is hard, and the margin of safety small." A family may want to nourish its babies but be unable to do so. As in many "primitive" societies, Eskimo mothers will nurse their infants over a much longer period of time than mothers in our culture. The child will take nourishment from its mother's breast for four years, perhaps even longer. So even in the best of times there are limits to the number of infants that one mother can sustain. Moreover, the Eskimos are a nomadic people—unable to farm, they must move about in search of food. Infants must be carried, and

a mother can carry only one baby in her parka as she travels and goes about her outdoor work. Other family members help whenever they can.

Infant girls are more readily disposed of because, first, in this society the males are the primary food providers—they are the hunters, according to the traditional division of labor—and it is obviously important to maintain a sufficient number of food providers. But there is an important second reason as well. Because the hunters suffer a high casualty rate, the adult men who die prematurely far outnumber the women who die early. Thus if male and female infants survived in equal numbers, the female adult population would greatly outnumber the male adult population. Examining the available statistics, one writer concluded that "were it not for female infanticide...there would be approximately one-and-a-half times as many females in the average Eskimo local group as there are food-producing males."

So among the Eskimos, infanticide does not signal a fundamentally different attitude toward children. Instead, it is a recognition that drastic measures are sometimes needed to ensure the family's survival. Even then, however, killing the baby is not the first option considered. Adoption is common; childless couples are especially happy to take a more fertile couple's "surplus." Killing is only the last resort. I emphasize this in order to show that the raw data of the anthropologists can be misleading; it can make the differences in values between cultures appear greater than they are. The Eskimos' values are not all that different from our values. It is only that life forces upon them choices that we do not have to make.

## 2.6 How All Cultures Have Some Values in Common

It should not be surprising that, despite appearances, the Eskimos are protective of their children. How could it be otherwise? How could a group survive that did not value its young? It is



easy to see that, in fact, all cultural groups must protect their infants:

- 1 Human infants are helpless and cannot survive if they are not given extensive care for a period of years.
- 2 Therefore, if a group did not care for its young, the young would not survive, and the older members of the group would not be replaced. After a while the group would die out.
- 3 Therefore, any cultural group that continues to exist must care for its young. Infants that are not cared for must be the exception rather than the rule.

Similar reasoning shows that other values must be more or less universal. Imagine what it would be like for a society to place no value at all on truth telling. When one person spoke to another, there would be no presumption at all that he was telling the truth for he could just as easily be speaking falsely. Within that society, there would be no reason to pay attention to what anyone says. (I ask you what time it is, and you say "Four o'clock." But there is no presumption that you are speaking truly; you could just as easily have said the first thing that came into your head. So I have no reason to pay attention to your answer; in fact, there was no point in my asking you in the first place.) Communication would then be extremely difficult, if not impossible. And because complex societies cannot exist without communication among their members, society would become impossible. It follows that in any complex society there must be a presumption in favor of truthfulness. There may of course be exceptions to this rule: There may be situations in which it is thought to be permissible to lie. Nevertheless, there will be exceptions to a rule that is in force in the society.

Here is one further example of the same type. Could a society exist in which there was no prohibition on murder? What would this be like? Suppose people were free to kill other people

at will, and no one thought there was anything wrong with it. In such a "society," no one could feel secure. Everyone would have to be constantly on guard. People who wanted to survive would have to avoid other people as much as possible. This would inevitably result in individuals trying to become as self-sufficient as possible—after all, associating with others would be dangerous. Society on any large scale would collapse. Of course, people might band together in smaller groups with others that they could trust not to harm them. But notice what this means: They would be forming smaller societies that did acknowledge a rule against murder: The prohibition of murder, then, is a necessary feature of all societies.

There is a general theoretical point here, namely, that *there are some moral rules that all societies will have in common, because those rules are necessary for society to exist*. The rules against lying and murder are two examples. And in fact, we do find these rules in force in all viable cultures. Cultures may differ in what they regard as legitimate exceptions to the rules, but this disagreement exists against a background of agreement on the larger issues. Therefore, it is a mistake to overestimate the amount of difference between cultures. Not every moral rule can vary from society to society.

## 2.7 Judging a Cultural Practice to Be Undesirable

In 1996, a 17-year-old girl named Fauziya Kassindja arrived at Newark International Airport and asked for asylum. She had fled her native country of Togo, a small west African nation, to escape what people there call excision.

Excision is a permanently disfiguring procedure that is sometimes called "female circumcision," although it bears little resemblance to the Jewish ritual. More commonly, at least in Western newspapers, it is referred to as "genital mutilation." According to the World Health Organization, the practice is

widespread in 26 African nations, and two million girls each year are "excised." In some instances, excision is part of an elaborate tribal ritual, performed in small traditional villages, and girls look forward to it because it signals their acceptance into the adult world. In other instances, the practice is carried out by families living in cities on young women who desperately resist.

Fauziya Kassindja was the youngest of five daughters in a devoutly Muslim family. Her father, who owned a successful trucking business, was opposed to excision, and he was able to defy the tradition because of his wealth. His first four daughters were married without being mutilated. But when Fauziya was 16, he suddenly died. Fauziya then came under the authority of his father, who arranged a marriage for her and prepared to have her excised. Fauziya was terrified, and her mother and oldest sister helped her to escape. Her mother, left without resources, eventually had to formally apologize and submit to the authority of the patriarch she had offended.

Meanwhile, in America, Fauziya was imprisoned for two years while the authorities decided what to do with her. She was finally granted asylum, but not before she became the center of a controversy about how foreigners should regard the cultural practices of other peoples. A series of articles in the New York Times encouraged the idea that excision is a barbaric practice that should be condemned. Other observers were reluctant to be so judgmental—live and let live, they said; after all, our practices probably seem just as strange to them.

Suppose we are inclined to say that excision is bad. Would we merely be applying the standards of our own culture? If Cultural Relativism is correct, that is all we can do, for there is no cultural-neutral moral standard to which we may appeal. Is that true?

## Is There a Culture-Neutral Standard of Right and Wrong?

There is, of course, a lot that can be said against the practice of excision. Excision is painful and it results in the permanent loss of sexual pleasure. Its short-term effects include hemorrhage, tetanus, and septicemia. Sometimes the woman dies. Longterm effects include chronic infection, scars that hinder walking, and continuing pain.

Why, then, has it become a widespread social practice? It is not easy to say. Excision has no obvious social benefits. Unlike Eskimo infanticide, it is not necessary for the group's survival. Nor is it a matter of religion. Excision is practiced by groups with various religions, including Islam and Christianity, neither of which commend it.

Nevertheless, a number of reasons are given in its defense. Women who are incapable of sexual pleasure are said to be less likely to be promiscuous; thus there will be fewer unwanted pregnancies in unmarried women. Moreover, wives for whom sex is only a duty are less likely to be unfaithful to their husbands; and because they will not be thinking about sex, they will be more attentive to the needs of their husbands and children. Husbands, for their part, are said to enjoy sex more with wives who have been excised. (The women's own lack of enjoyment is said to be unimportant.) Men will not want unexcised women, as they are unclean and immature. And above all, it has been done since antiquity, and we may not change the ancient ways.

It would be easy, and perhaps a bit arrogant, to ridicule these arguments. But we may notice an important feature of this whole line of reasoning: it attempts to justify excision by showing that excision is beneficial—men, women, and their families are all said to be better off when women are excised. Thus we might approach this reasoning, and excision itself, by asking which is true: Is excision, on the whole, helpful or harm-

ful?

Here, then, is the standard that might most reasonably be used in thinking about excision: We may ask whether the practice promotes or hinders the welfare of the people whose lives are affected by it. And, as a corollary, we may ask if there is an alternative set of social arrangements that would do a better job of promoting their welfare. If so, we may conclude that the existing practice is deficient.

But this looks like just the sort of independent moral standard that Cultural Relativism says cannot exist. It is a single standard that may be brought to bear in judging the practices of any culture, at any time, including our own. Of course, people will not usually see this principle as being "brought in from the outside" to judge them, because, like the rules against lying and homicide, the welfare of its members is a value internal to all viable cultures.

## **Why Thoughtful People May Nevertheless Be Reluctant to Criticize Other Cultures.**

Although they are personally horrified by excision, many thoughtful people are reluctant to say it is wrong, for at least three reasons. First, there is an understandable nervousness about "interfering in the social customs of other peoples." Europeans and their cultural descendents in America have a shabby history of destroying native cultures in the name of Christianity and Enlightenment, not to mention self-interest. Recoiling from this record, some people refuse to make any negative judgments about other cultures, especially cultures that resemble those that have been wronged in the past. We should notice, however, that there is a difference between (a) judging a cultural practice to be morally deficient and (b) thinking that we should announce the fact, conduct a campaign, apply diplomatic pressure, or send in the army to do something about it. The first is just a matter of trying to see the world clearly, from a moral

point of view. The second is another matter altogether. Sometimes it may be right to "do something about it," but often it will not be.

People also feel, rightly enough, that they should be tolerant of other cultures. Tolerance is, no doubt, a virtue—a tolerant person is willing to live in peaceful cooperation with those who see things differently. But there is nothing in the nature of tolerance that requires you to say that all beliefs, all religions, and all social practices are equally admirable. On the contrary, if you did not think that some were better than others, there would be nothing for you to tolerate.

Finally, people may be reluctant to judge because they do not want to express contempt for the society being criticized. But again, this is misguided: To condemn a particular practice is not to say that the culture is on the whole contemptible or that it is generally inferior to any other culture, including one's own. It could have many admirable features. In fact, we should expect this to be true of most human societies—they are mixes of good and bad practices. Excision happens to be one of the bad ones

# *Part 15: What is Moral Relativism?*

## Going Meta

You might have heard the term ‘meta’ before. For example, when I was in undergrad, I was explaining my Epistemology course to some friends, saying something like “it’s the study of knowledge, we are learning about knowledge”, and one person in the group said “that’s so meta!”. The term ‘meta’ comes from Greek, meaning ‘over’, and has there a similar sense to our prefix ‘trans-’ (as in ‘translate’ or ‘transfer’), which comes from Latin and has the same sense as the Greek. In English, also, there are two senses of ‘over’, the first means ‘across’, like to carry something over to another, and the second means ‘above’, as in to put something above another. ‘Meta-’ and ‘trans-’ differ in the same way. With very few exceptions (‘metaphor’, which comes from the Greek for ‘carry over’), the prefix ‘meta-’ means that you are doing something above, over, the area in question and the prefix ‘trans-’ means that you are doing something across, over, certain boundaries. In the case of ‘translate’, this comes from the Latin verb ‘transfero’, meaning “I carry over”, that verb is also where we get the English verb ‘transfer’.

More explicitly, when **meta-** is attached to a field of study,

it refers to a field of study one step more abstract than it. One way to think about this is that I am now asking questions about the questions in the field. This is not just true for fields in philosophy, though once the level gets high enough, it becomes an area of philosophy, but it's also true for any science or subject you pick. Here is a table of some examples:

| Subject     | Questions                                             | Meta-Subject    | Meta-Questions                                                                    |
|-------------|-------------------------------------------------------|-----------------|-----------------------------------------------------------------------------------|
| Biology     | Concerning living organisms                           | Metabiology     | What exactly is living?<br>What exactly is an organism?                           |
| Physics     | Concerning matter and its actions in space and time   | Metaphysics     | Is matter all that exists?<br>What is space?<br>What is time?                     |
| Linguistics | Concerning language, meaning, and structure in humans | Metalinguistics | How does meaning relate to truth and intentions?<br>How do words pick out things? |

When you 'go meta', you are taking yourself out of the sphere of the field and starting to ask questions about the very nature of it. Due to the nature of philosophy and philosophers in general, we are far more inclined to go meta on a topic. So, for any field in philosophy, there's likely a well-trodden path for the various meta-questions and sometimes a well-trodden path for the meta-meta-questions.

The various stances which can be generated once you choose to go meta on a topic tend to fall into nice, neat, easy to classify packets. The first level has two different kinds of stances, **Realism** and Anti-Realism (sometimes called **Nihilism**, which is what I prefer). Realism claims that the facts about the subject are real, hence the name, while the Anti-Realist claims that



they aren't real. And then, depending on what the stance says about various meta-level questions, they are further subdivided. In the Realist side, there are two general stances which get our attention, Relativism and Objectivism.

There are different kinds of Relativisms and Objectivisms (the typical subdivisions of Realism) out there, and those tend to be more particular to the field and there may be different kinds of Nihilisms aside from Error Theory and Expressivism, but those go well beyond the scope of this class. Towards the end of this module, I will explain those stances in the Nihilism branch for Meta-Ethics in more detail.

## The Start of Moral Relativism

Moral Relativism is a part of a family of stances, all of which say that the truth of something, or everything, is relative (generally called **Relativism**). What is relative or to what it is relative all depends on the stance. This sort of idea is found in several areas of philosophy and the notion (applied differently) is found in physics. This, in its broad sense, is the stance that there is no absolute truth (objective truth), but rather the truth about things varies from person to person or from culture to culture, depending on the version of relativism which is held.

Relativism comes in many different flavors. The first is that everything is relative, this is global relativism. The other is that only certain things are relative, this is limited relativism. From this, we have to ask what are they relative to? These are either to your culture or to you individually.

**moral relativisms** is a limited form of relativism. It does come in two flavors, Moral Individual Relativism (what is moral is relative to the doer, and there is no way for another person to say that what they did is/was morally wrong) and Moral Cultural Relativism (what is moral is relative to the culture and there is no way of saying that they are wrong from an outsider's perspective). One can also say that Moral Relativism is

the stance that there is no objective basis for claiming that a particular action, or a class of actions is right or wrong, permissible or impermissible. Or, that there is no fact of the matter regarding whether actions are right or wrong. Whether something is, in fact, right or wrong is relative to the culture or to the person. Some classic examples of this are killing babies, cannibalism, and funeral rights.

## The Cultural Difference Argument

Often, people argue for moral relativism by pointing to examples, it is often claimed that due to the differences in moral beliefs, there must not be an objective, absolute morality. If you remember from my charts, the moral relativism still holds that there are moral truths, but those truths are relative to the culture. Cultural Moral Relativism is by far the most common stance found in popular culture, so that's the one we will point to, but individual moral relativism has similar problems as well as others unique to it. Here is a classical example used to argue for Cultural Moral Relativism:

Darius, a king of ancient Persia, was intrigued by the variety of cultures he encountered in his travels. The Callatians (a tribe of Indians) customarily ate the bodies of their dead fathers. The Greeks practiced cremation. Darius thought that an understanding of the world must include an appreciation of such differences. One day, he brought some Greeks who happened to be present and asked them what they would take to eat the bodies of their dead fathers. They were shocked and replied that no amount of money could persuade them to do such a thing. Then Darius called in some Callatians, and while the Greeks listened asked them what they would take to burn their dead fathers' bodies. The Callatians were horrified and told Darius not even to mention such a dreadful thing. The two people had a very similar reaction to the opposite actions.<sup>a</sup>

---

<sup>a</sup>Herodotus, *The Histories*, translated by Aubrey de Sélincourt (Penguin Books, 1988).

From this, we can see that two cultures, when they encountered the customs of the other were equally mortified. But sometimes, even the most 'socially aware' of us will be mortified by customs found in other cultures. Take this example:

In some Inuit cultures, there is a common practice known as 'the dip' (my translation). When a child is born in the dead of winter, it is common for the mother to carve out a hole in the ice and place the child in the water. This kills the child instantly (infanticide). This is done purely at the parents' discretion and there is no negative stigma about it. Old people also, when they became too feeble to contribute to the family, were left out in the snow to die.

The lesson from these two stories (the second of which I

have seen it used as examples enough for me to think it's true) is that different cultures have different moral codes. What is seen as just fine in one culture is seen as horrifying in another. For a fictional case, take this example:

For most of the world, the elderly are treated with respect and cared for until their natural death. The Kaelon people are different. In this society, the prevailing view is that it's the duty of the 'elderly' to leave their tasks to the next generation, that forcing the next generation to care for the elderly is cruel, and that having death come for a person seemingly randomly is heartless. So, when a Kaelon turns 60 years old, they undergo the Resolution. In it, a great party is thrown, celebrating their life and accomplishments, and afterwards, the Kaelon commits suicide. Living past this point is seen as greedy, their time is up and their accomplishments after this point are seen as stolen from the next generation. This is expected of all Kaelons, refusing to kill yourself at the allotted time will cause even family members to be ashamed of you. If a Kaelon seeks asylum to avoid the Resolution, the Kaelons will declare war on the other culture in order to kill the person.<sup>a</sup>

---

<sup>a</sup> "Half a Life," writer Ted Roberts and Peter Allan Fields, director Les Landau, 1991, Paramount Domestic Television.

Many have looked at examples like this and have thought that there must not be an actual objective truth about morality, that it must merely be opinion, either held by a person or collectively as a group. This leads us to:

## The Cultural Differences argument

These examples, and my not so subtle hints, point to a very interesting argument. It seems, given how I have framed it,

the most socially conscious stance would be one which puts all cultural customs on equal footing, to not have some objective measure which can be used to discriminate against a culture.

So, using this observation, we get this argument:

|                                                          |                                                                    |
|----------------------------------------------------------|--------------------------------------------------------------------|
| Different cultures have different moral codes.           | There is disagreement about morality.                              |
| Therefore, there is no objective "truth" about morality. | Therefore, there's no objective fact of the matter about morality. |

But, does this show that the next module in this class is going to be gibberish? Well, no. Though there are some moral relativists out there in philosophy today, there are very few, and no philosopher (no really good philosopher) is a moral relativist because of this argument. There are many problems with it which are worth exploring.

### Problem 1: It isn't a good argument

For an argument to be good, philosophers have a pretty high standard. The conclusion needs to follow from the premises necessarily. There can be no case where the premises are true and the conclusion false (this is called 'validity'). If an argument is not valid, then it is not sound. The first line of the argument concerns what people believe while the conclusion concerns what is actually the case. The issue with this sort of reasoning can be seen using a similar sort of method as the Limitation of the Conclusion problems in the arguments for the existence of God. If you think that The Cultural Differences Argument is accurate, then you will, by the same reasoning, need to think that these arguments are accurate:

|                                                                                |                                                                |
|--------------------------------------------------------------------------------|----------------------------------------------------------------|
| Different cultures/people have different beliefs about the shape of the earth. | Different people have different beliefs about global warming.  |
| Therefore, there is no fact of the matter about the shape of the earth.        | Therefore, there's no fact of the matter about global warming. |

This argument follows the exact same reasoning as the cultural differences argument. If it works for one then it should work for the other. But, we can also look at the spirit of the argument, and attack that as well.

## **Problem 2: Cultural Oppression**

Though the argument may not be any good, all that shows is that the argument isn't good. It does not show that morality is objective. But, if we look at the last example I gave, involving the Kaelons, we see something interesting. If morality is relative to a culture, there's nothing stopping one culture, morally speaking, from oppressing people in their culture and oppressing people in other cultures. For example, if there are no objective moral facts, then there would be nothing wrong with the Kaelons declaring war on another culture for allowing someone to receive asylum. Similarly, there would be nothing wrong with the Kaelons forcing members of their culture to unwillingly commit suicide. This sort of case can be found in history as well, with the sheer number of examples being too numerous to list. This is a problem which will appear again later in this class, because it's more obvious inside another argument.

## **The Cultural Imperialism Argument**

The Cultural Differences Argument is not the only argument in favor of Cultural Moral Relativism, there are several others. Most fall to the same sort of objections which we will see later, others have problems of their own. The Cultural Imperialism Argument is a little more advanced, claims to prove this sort of relativism, but it, equally, has issues. Here is the introduction:

### **Cultural Imperialism**

Throughout history, we have seen cases of one culture forcing beliefs and moral stances onto another. For example, we have

the forced religious conversions of the Aztecs by the Spanish, we have the boarding schools which the US put Native American children in, and in China, we have the Uyghurs being put into 'reeducation camps' (this may still be happening, this is true as of 2018). All of these examples have/had the same purpose: changing the practices, religions, and moral ideologies of the people.

Cultural Imperialism is the reason for these behaviors, why cultures force themselves on others in this way (and sometimes in even more violent ways). Cultural Imperialism is the stance your culture as all of the right answers, morally speaking, your culture is perfect, and because of this, you are justified in forcing other cultures to change (typically to your own). But, looking at the number of examples from history, we can see that this imperialism is wrong, it should not be done. This can be seen both from an insider's and an outsider's perspective. Naturally, this perspective on Imperialism leads easily to Cultural Moral Relativism, because it seems never to be right to judge the values of one culture against the values of another. Put more formally, we get this argument:

### **The Cultural Imperialism Argument**

1. Cultural Imperialism is morally wrong. (Intuition).
2. This follows from the examples which I have given, many argue that this sort of behavior is wrong and should not be done at all. The main reason for the Prime Directive in Star Trek is because of the base line principle, you need to let cultures' internal matters be internal matters.
3. If Cultural Imperialism is morally wrong, then it is wrong to judge the values of one culture against the values of another. (Consequence of it being wrong)
4. This, in turn, is a consequence of the wrongness of Cultural Imperialism (the first line). Basically, if it was OK

to make judgments about other cultures' practices, then it would be OK to act on those judgments. Acting on those judgments would be Cultural Imperialism, which we said was wrong.

5. If it is wrong to judge the values of one culture against the values of another, then no person is ever justified in criticizing the moral norms of another culture. (Consequence of the consequence)
6. This is a consequence of the wrongness of the judgments about a culture. Criticizing a culture's practices, in either word or action, is wrong by the same reasoning as Cultural Imperialism. The exact reasoning is found in the previous line.
7. If no person is ever justified in criticizing the moral norms of another culture, there are no non-relative moral truths (they are all relative to culture). (Consequence)
8. This is the final line of support for the conclusion. If there were an absolute, objective, moral rule, or set of rules, then it would be possible for a person to be correctly justified in criticizing another culture's norms. All they would need to do is point to the objective rules and show that the culture is in violation of them. But, we have shown that a person can never be justified in this way, so, it would seem that there must not be an absolute moral standard (they are relative to the culture).
9. Therefore, there are no non-relative moral truths (they are all relative to culture).

Now, despite my build-up and explanation for this, there is a major issue with this argument, both in the way it's constructed (though, technically valid) and the spirit behind it. For example, I can use this argument, exact same premises, to show that unicorns exist,  $2+2=5$ , and anything else I could want.



## The Problem

As I mentioned above, I can use this argument, not the structure, but the argument itself, to prove that mermaids exist. Any argument with this capability will be valid, but it will not be sound, there will be a factual error in it somewhere. For this one, let's try and work out the factual error, and here it's fairly easy to spot (hint, look at the first line and the last line):

|                         |  |                           |
|-------------------------|--|---------------------------|
| Cultural Imperialism is |  | There are no non-relative |
| morally wrong.          |  | moral truths.             |

The problem with this argument is that it's contradictory. You are saying that the same is both true and false. In the first line, we are saying that Cultural Imperialism is wrong, which is not contradictory on its own (if you think that it can be OK, then you have included other aspects into it). Similarly, the last line says that all moral truths are relative. Again, not contradictory on its own. But, in combination, the lines conflict. The first line is a non-relative moral truth and the last line says that those don't exist. This also attacks the spirit of the argument, in that the core premises/intuition which build it up come from the idea that Cultural Imperialism is wrong, which contradicts the point it's trying to prove.

There are a few ways out for the Relativist. First, they could accept that some cases of Cultural Imperialism are fine, making it not always wrong. For example, we can see that the Allies entering WW2 to stop certain behaviors was Cultural Imperialism, but that wasn't wrong. The issue with this is that if we go down this rabbit hole, we find that the cases where Cultural Imperialism is permissible are cases of objective moral truths, which limits the initial scope of Objectivism, but contradicts Relativism.

Another way out, which limits Cultural Moral Relativism further, is to say that there's a distinction between internal and external behaviors. Internal behaviors are those which only apply to members of the same culture, while external behaviors are those which involve members of different cultures. Cultural

Imperialism is an external behavior. The Relativist here can claim that the morality of internal behaviors is relative while the morality of external behaviors is objective. This has its own issues as well, mostly dealing with cases like the Kaelons which I gave before, and can be easily seen in that case.

## Tests to Tell Whether You are *actually* a Relativist

These are some tests which you can preform on yourself or on others, or on groups, which will tell you your actual feelings about moral relativism. There are three tests total, sometimes these tests will work individually, sometimes jointly (depending on the case), but they can be very useful. These are based on *Why I am an Objectivist about Ethics (And Why You Are, Too)* by David Enoch.<sup>118</sup>

### The First Test: The Spinach Test

This is a pretty straight forward test for whether you actually are a moral relativist. But, as a fair warning, philosophy kills jokes. As an example, and where the test gets its name, lets look at the following joke:

A kid hates spinach, later, he says that he's glad he hates spinach. When asked why, he says "because if I liked it, I would eat it; and it's yucky!"

More often than not, we will take this as funny. And, in a lecture format, if I present it right, I will get a laugh. But, the test is not whether or not you laugh at this joke, but rather whether you laugh at jokes of this form/structure, with different

---

<sup>118</sup>David Enoch, "Why I Am an Objectivist About Ethics (And Why You Are, Too)," *The Ethical Life*, 3rd ed. Edited by Russ Shafer Landau (Oxford UP, 2014).

things put in the place of 'spinach'. For example, let's try it with entertainment and other foods:

A kid hates watching golf. He says that he's glad he hates watching golf because "if I liked it, I would watch it, and it's boring!"

A metal-head says that he's glad he hates country. When asked why, he replies 'because if I liked country, I would listen to country, and that's just bad music.'

George H.W. Bush hates broccoli and says that he's glad he hates broccoli. The reason? "Because if I liked broccoli, I would eat it, and it's gross."

These are the same, if presented right, they will get a laugh. But, will this joke work for anything which I plug into the joke's formula? What if I plug in something which is more absolute? Take these examples:

A 20th century man believes that the Earth is round. He claims that he's glad that he's a 20th century man because "If I grew up in the first century, I would have believed that the Earth is flat, and that's just wrong."

A 20th century man believes that women should have the right to vote. He claims that he's glad that he's a 20th century man because "If I grew up in the 19th century, I would have believed that women shouldn't vote, and this is wrong."

An Iraqi woman who grew up in the 2000s believes that women are not equal to men. She claims that she's glad about this because "if I grew up in the 60s and 70s, I would have believed that we were equal, and this is wrong."

Something is very different here. If I tried to present these as a joke, even with the same tones and inflections, I would not get a laugh (and I don't, I have tried). But, what's the

difference between these? What sort of things get a laugh and what sort of things don't? Well, the difference between these cases is seemingly unrelated, but it's the only one which works (they are related). The relative things, the things which are merely preference, those get a laugh when plugged into the joke, the things which aren't relative, the objective things, those don't get a laugh. So, this test is basically if you get a laugh, it's relative, if you don't, then it's objective. The vast, vast majority of the time, we do not laugh about morality. So, we, at least at some level, think that morality is objective.

### **The Second Test: The Disagreement Test**

Often, in real life, we get into disagreements with people. Sometimes these are serious, like whether refugees should be allowed in the country or whether global warming is occurring. Other times they are just silly, like whether broccoli is yucky or whether dark chocolate is better than milk. The second test for relativism about a subject concerns the nature of such disagreements. For this test, you need to imagine, or actually get into, a disagreement about something. Ask yourself what the disagreement feels like. For example, take these examples:

|                                                                         |                                                                        |
|-------------------------------------------------------------------------|------------------------------------------------------------------------|
| A disagreement about whether the Earth is flat.                         | A disagreement about a mathematical principle.                         |
| A disagreement about whether refugees should be allowed in the country. | A disagreement about whether slavery was bad.                          |
| A disagreement about whether God exists.                                | A disagreement about what to do in the trolley problem is permissible. |
| A disagreement about whether pineapple should go on pizza.              | A disagreement about how enjoyable Buffy the Vampire Slayer was.       |
| A disagreement about the best Metal band.                               | A disagreement about the direction of a toilet paper roll.             |

These disagreements are very different. Think about what it's like to be in those kinds of disagreements. In the case of broccoli or chocolate, you are trying to state your own preference and then, maybe, trying to get the listener to come to their senses and change theirs. However, when we are engaged in a disagreement about, say, global warming, the point of the argument is not to change the other person's mind or to state how you feel. Rather, it is to get at the truth. Looking above at the examples, do the ones in the left column feel the same? Do they all feel like facts? Similarly, do the ones in the right column all feel the same? Do they feel like preferences? If the feeling of the disagreements in the left column all feel the same, then you are an objectivist about morality, in your heart.

So, for this test about whether you are a relativist about some subject, you need to ask yourself what it feels like to be in a disagreement about the subject. Imagine that you are having a disagreement with someone about the morality of abortion. How does that feel? Does it feel like you are having a discussion about pineapple and green-olive pizza? Or does it feel like you are having a discussion about global warming?

Yet again, very rarely do we feel like disagreements about morality are disagreements about preference. When we have those debates, it feels like they are about facts.

### **The Third Test: The What-If Test**

Like the other two, this one is a bit different. It involves what are called counterfactuals. They are called this because they are counter to the facts. We use these sort of statements all the time without realizing it. Any time you ask yourself what would happen if something were the case, you are using a counterfactual. These are insanely useful in philosophy as well as in science. There's an entire cottage industry in philosophy dedicated to coming up with how we use, accept, and reject these kinds of statements. That being said, since we use them so much, your gut intuition will suffice. For this test, we will present the subjects as yes-no 'what if' style questions. Such as, 'what if I drove a 110MPH on a 60MPH highway, would that be dangerous?' It is not true that I drive that fast, but I am trying to figure out what would be the case if I did. Here are some examples of what-if style questions which show this test at work.

If the vast majority of people believed that politicians were actually aliens, would they still be human?

If the vast majority of people believed that global warming wasn't happening, would it still?

If the vast majority of people thought that gender discrimination was fine, would it still be wrong?

If the vast majority of people thought that the Earth was flat, would it still be round?

If the vast majority of people thought that Philosophy was pointless, would it still be valuable?

If the vast majority of people thought that pineapple on pizza was delicious, would it still be gross?

If the vast majority of people wore top hats and thought they were cool, would they still be out of fashion?

If the vast majority of people thought that Buffy was the worst show ever, would it still be great?

If the vast majority of people thought that sweet potatoes were tasty, would they still be gross?

If the vast majority of people thought that spinach was gross, would it still be tasty?

For each of these, there should be a yes or no answer, one which you can easily give. In some versions of this, I have presented it as "all people" rather than "the vast majority of people", but that has led to some confusion. As before, the normal responses to the questions in the left column are all 'yes', while the answers to the questions in the right are normally 'no'. This test, if the question is formulated correctly, is the most definitive, I think. It points out the fact that objective truths are not up-to-us, they will be the case regardless of whether or not people discover them.

The steps to preform this test on yourself are pretty straight forward. First, take a stance about something which you think is true. For example, you can have that slavery is bad, that pineapple on pizza is good, that the Kaelons shouldn't force

their elderly to commit suicide, that the Earth is round, or something like that. Next, you imagine a case where the vast majority of people believe the opposite. And, finally, ask yourself whether, in this strange world, you would still be correct. If the question is phrased correctly (it's possible to phrase them so that you will get, consistently, the opposite answer) and if you get the answer 'yes', then the subject matter is objective and if you get the answer 'no', then the subject is relative.

## Consequences of Taking Moral Relativism Seriously

So far, we have seen evidence that we don't actually think that moral relativism is correct and we have seen that the arguments for the stance aren't any good (it's possible that there are others, and likely are others, but they will all, likely, fall you the same issues). But, like with the arguments for the existence of God, an objection to the argument doesn't show that it's conclusion is false. There are some true or false things which just can't be proven one way or the other (for example, "the first digit of TREE(3) is 1" is either true or false, but due to the size of the number, it's just impossible for us to prove one way or the other). Now, to actually disprove Cultural Moral Relativism we need to attack it head on, not just the supporting evidence. To do this, we will use a method which you might recognize from science (science got it from Philosophy). We will assume that relativism is true and show that it does not line up with the world, or that the consequences of it are so outlandish we can dismiss it easily. There are, at the heart of Moral Relativism, 3 problems which we will address (and more, but 3 is enough).

### The Criticism Problem

If we remember back to the Cultural Imperialism Argument, one of the core reasons it works is because we want to say that



criticism of other societies (through word or action) is morally wrong. Though this does contradict Cultural Moral Relativism, that stance does have a variant to it built in. Namely:

We could no longer say that the customs of other societies are  
morally inferior to our own

In other words, if moral relativism is correct, we can't say, truthfully/correctly, that any culture has 'got it wrong', we can't say, in that way, that any cultural behavior is wrong. Sometimes, this seems fine, like in cases where doing so leads to the wrong sort of Cultural Imperialism. Other times, however, it leads to some very outlandish conclusions like:

We can't say that a culture torturing and/or killing other groups systematically is morally wrong.

We can't say that intolerance is bad.

We can't say that cultures which engage in hateful acts towards others are morally wrong.

We can't say that tolerance is good.

The argument for this conclusion rests on the idea that if we accept moral relativism, there would be no gauge, objective standard, to measure morality.

1. If MCR is true, then morality is relative to culture. (This is basically the definition)
2. If morality is relative to the culture, then there is no standard to gauge the morality of a culture.
3. If there is no such standard, there is no way to compare one culture to another (morally speaking)
4. If there is no way to compare them, then there is no way to say that the Nazis were bad.
5. So, if MCR is true, then there is no way to say that the Nazis were bad.

6. But, the Nazis were bad!
7. Therefore Moral Cultural Relativism is false.

The way out, for the relativist, is to claim that the Nazis and every other culture which has engaged in horrible acts weren't bad. This can be quite the pill to swallow.

## The Sheeple Problem

Some of you may have heard the term 'sheeple' before. It's a mash-up of the word 'sheep' and the word 'people'. In this case, if we take Moral Cultural Relativism seriously, thinking for yourself about morality is always morally wrong. This is not directly because of some universal moral rule, which would contradict Moral Relativism, rather it's because if something is a moral rule, then not abiding by it is morally wrong, and since every culture will have moral rules, not blindly following them will be wrong. Rather than speaking in negatives, we have this statement:

We could decide whether actions are right or wrong just by consulting the standards of our society.

Moral Cultural Relativism gives us a very simple test to tell whether an act is right or wrong. Just look at the cultural standard. Also, at the same point, trying to change those standards, whatever they may be, is breaking those standards, and thereby wrong. This means that, for example: Suppose that a person in the deep south (Pre-Civil War) was curious about whether slavery was permissible. All they would need to do is ask whether it fit with the code of the society. If it did, then they would be OK with having slaves. Another way to think about this is to imagine that God, or whoever, once the criteria for being a culture are met, sends down a book with the rules. Doing anything other than blindly following those rules is wrong, regardless of the culture. This means that every

civil rights movement in history, in every culture, was morally wrong. This includes Black Rights Movements and Women's Rights Movements.

But, no cultural system is ever perfect and they have room to improve. Not only can't we criticize other cultural codes, but we can't criticize our own. However, claiming that every civil rights movement in history was morally wrong is absolutely outrageous, so Moral Cultural Relativism must be incorrect.

This leads us to the following argument:

1. If MCR is true, then morality is relative to culture. (This is basically the definition)
2. If morality is relative to the culture, then whether an action is right or wrong is determined by your cultural norms.
3. If those are determined by the norms, then trying to change those norms is always morally wrong.
4. So, if MCR is true, then trying to change moral norms is always morally wrong.
5. But, trying to change moral norms is sometimes morally right.
6. Therefore, Moral Cultural Relativism is false.

The way out for the relativist here is to claim that moral change is always wrong. Basically, claim that Martin Luther King Jr. was about as immoral as they come.

## The Progress Problem

Many of us, especially some of the more 'woke' members of our society, will claim that some moral change is for the better. But, the question quickly becomes "what is 'better'?" When we use this term, we are comparing one thing to another. We say

that "pineapple and green-olive pizza is better than pepperoni" or "Pushing Daisies is better than The Big Bang Theory" or "carrots are better than potatoes", in the context of healthiness. But, when we say that some moral change is for the better, what are we comparing and what are we using to make the comparison?

When we claim that some moral change is for the better, we are comparing where we were, morally, and where we are, morally. To make this comparison, if it's correct, we are using an absolute moral standard, a universal standard. We mean that we have moved closer to getting morality right. We call this progress. If we take Moral Cultural Relativism seriously, there's no moral standard, there's nothing to use to compare where we are and where we were morally speaking. We can't, correctly, claim that there's progress.


The idea of moral progress is called into doubt

This means that not only can't we compare other cultures, but we can't even compare the same culture to itself at a different time. This leads us to the following argument:

1. If MCR is true, then morality is relative to culture. (This is basically the definition)
2. If morality is relative to the culture, then there is no standard to the morality of a culture.
3. If there is no such standard, there is no way to compare one culture to another (morally speaking)
4. Different people at different times, if the norms changed, are, for all intents and purposes, different cultures.
5. So (from 1, 2, 3, and 4), if MCR is true, then there is no way to compare people from one time to another (morally).

6. If there is no way to compare people from different times, there can be no way for us to say that we are better than we were (made progress).
7. So, if MCR is true, then there is no way to say that we have made progress.
8. But, we have progressed (at least in some areas).
9. Therefore, Moral Cultural Relativism is false.

The way out for the relativist is to claim that we haven't made progress, in any way. We are no better than we ever were.



## *Part 16: So, If Moral Relativism is Wrong, What Next?*

If you are willing to say that moral cultural relativism is wrong, and, given the general responses to the various problems with moral relativism as well as what can be expected from the tests, you likely are willing to say that it's wrong, we have some problems. Namely, what is moral? Who decides who's correct in a moral conflict? Where do moral truths come from? Here I will cover some of the basic questions which former relativists tend to give and the general responses (sometimes, the response will need to be more particular).

### **If culture's don't choose, what is morality?**

There are a few ways to go about replying to this question. Morality/ethics centers around 'should' questions of a certain kind. When we ask questions like these, for example "what should I do?", "should you call a doctor?", "should you get the

assignment in on time?”, there are two different senses, which may or may not overlap, depending on context and relevance. Something is moral/ethical when it’s the correct answer to the question ”what should I do?”, when we aren’t talking about practical cases (a practical case is one where you are asking about the correct way to perform some task, like changing the oil on your car). The cultural moral relativist would claim that the correct answer to the question in non-practical cases (for example, should I flip the switch in the trolley problem?) will depend on your culture and the norms associated with it. The moral objectivist, a moral realist who is also a non-relativist, would claim that there’s an actual answer to this question, which is not determined by your culture. The next module of this class covers some of the ways which philosophers have answered how to answer that question, without being relativist. Ethics is, at it’s core, trying to find the correct answer to this question.

You will notice that when I give examples of cultural norms, most of the time, these are cases where the culture got the answer wrong. Ethical theories are basically hypotheses about the correct way to get the answer to the question. Some theories point to absolute, objective duties which a person has (typically, it’s wrong to be irrational; acting in a certain way is irrational, therefore those ways of acting are wrong). Others point to the well-being of those affected (if an action makes the affected better off than otherwise, that’s the right action). Others still point to some exemplar of morality, some person who has the perfect character and then asks ”what would that person do?”. You may have heard that kind of thinking before with ’WWJD’.

## Who decides who’s right in moral conflict?

A moral conflict is a case where two or more people/groups disagree about the answer to the ”What should I do?” question,

in the relevant sense. The moral cultural relativist has a simple way to answer it, almost too simple, "if it's two groups in the same culture, the cultural norms settle the conflict, if it's two cultures, there's no real conflict." But this just does not seem right, if you recall the last page, were this true, all civil rights movements would wrong, because the cultural norms would settle the conflict in favor of the oppressing group. Very few of us would want this. So, that's out, but is there a non-cultural way to get the answer? This is the quest for the moral realist. It's not going to be based on belief (as the relativist points to).

Recall the Cultural Differences Argument. This argument relies on moral conflicts/disagreements. But, do we always settle debates like that? There may be disagreement about something, anything (for that matter), but it doesn't follow that there's no fact of the matter. If there's a conflict about math, we don't say that there's no answer, we consult the rules of mathematics (as they were discovered, not invented). When there's a conflict in science, we don't say that there's no answer, we perform experiments and discover the truth. For the moral realist, settling moral conflicts is more like settling a conflict in math or science than one in art. Realists perform experiments and consult the rules of morality to settle the debates (and the experiments either further support the rules or give examples of amendments which should be made). The quest to get the non-relativist rule to settle moral conflicts is hard, but it does not seem impossible, we do make progress in it. There's even a scientific-style method for figuring out which ethical theories should be applied and/or how they should be amended (as we will see in the next module).

In short, no one decides who's right in a moral conflict, it's just a fact that one is right and the other is wrong.



## Where do moral truths come from?

Moral truths are the answers to the "what should I do?" question (in the relevant sense). The moral cultural relativist has a simple, again almost too simple, answer "from the beliefs of the culture." Yet again, how can we possibly say that a culture has the wrong answer to the question when the culture chooses the answer? Like the issues before, this just can't be right. So, the question for the realist is just that, where do they come from?

For the realist, this is like asking "where do scientific truths come from?" or "where do mathematical truths come from?". Some moral realists point to God and say that all truths flow through Him, moral truths included, saying that God made them (an easy way to get all-knowing and all-good). Others point to the sort of creatures we are and our place in nature, these give us moral truths. Others still point to abstract notions of well-being or rationality. And even more say that moral truths are things which just could not have been otherwise, they just always are, they didn't "come from" anything. There is some debate about where they come from.

But, don't lose heart! There's just as much debate about where mathematical and scientific truths come from. For example, in the case of math, we could claim that the truths are necessary, they could not have been otherwise. We could claim, with just as much evidence, that the truths are constructs of the definitions of the terms which we are using overlapping in consistent ways ( $1+2=3$  because of the definitions of the terms and operations). And we could also claim that the truths came from God. As it turns out, in philosophy, the most common stance (for where mathematical truths come from and where moral truths come from) is that they always were and always will be.

## What about forcing them onto another culture?

This is a hard problem, because one of the core intuitions behind the Moral Relativist, and one of the big reasons why the stance is contradictory, is that it's wrong to force your morals onto another. The Moral Objectivist will claim that under the right circumstances, it's morally required that you force another to change. To some, this feels wrong. This feeling, often, gets traced back to the horrors of history, which I used as examples for the intuition behind the Cultural Imperialism Argument, where one group imposed themselves on another. At the same point, we need to look at history as a whole. There have been times when it was actually correct for one group to come in and force a change. In those cases, the imposing groups had real morality on their side and in the cases where it was wrong, we can say that they were mistaken about morality. For example, the Nazis in WW2 were just wrong about morality and the groups which came in were right. So, the Moral Objectivist just needs to be sure that their moral theory is getting the right answer in the case. This is an epistemological worry for the objectivist, not an ethical, meta-ethical, or metaphysical one.

Interestingly, when we look at our own history, we find things which we are ashamed of, things which we condemn the previous generation for. How and when this is correct is easy for us to see. Since different people at different times are, basically, different cultures, the exact same, or a similar, thought process applies when condemning or appreciating the behaviors of another culture. Kwame Appiah in *What Will Future Generations Condemn Us For?*<sup>119</sup> gives us some compelling tests for when future generations will be ashamed of our practices.

---

<sup>119</sup>Appiah, Kwame Anthony. "What will future generations condemn us for?" 26 Sept. 2010, [www.washingtonpost.com/wp-dyn/content/article/2010/09/24/AR2010092404113.html?hpid=opinionsbox1](http://www.washingtonpost.com/wp-dyn/content/article/2010/09/24/AR2010092404113.html?hpid=opinionsbox1).

These same tests apply to the behaviors of other cultures and tell us when it may be OK to force a change.

## Moral Nihilism

Moral Relativism and Moral Objectivism (Realism) are not the only meta-ethical stances, there are two others worth noting (though this does go beyond the necessary scope of this class). Like Moral Relativism, these stances deny that there are absolute moral truths. But, unlike Relativism, these theories claim that they aren't relative. They claim that there's no truth to them at all, regardless of context. Theories which do this fall into a category called 'Moral Nihilism'. Like Relativism, Nihilism comes as a family and you can pick and choose accordingly. Skepticism is most like Nihilism in how the family tree is organized.

Moral Nihilism, like Moral Relativism, is limited, it only makes claims about moral claims. Like Relativism, if you find one absolute, objective truth, then Global Nihilism is false, but for Limited Nihilism, you would need to find it in the context it's limited to. The two major theories which call Moral Nihilism home are 'Error Theory' and 'Expressivism'. Of the two, Error Theory is the strongest, but it can be quite counter-intuitive.

## Error Theory

Error Theory, applied differently, is a concept and stance which can be found in other areas of philosophy, though I have personally mostly encountered it in Metaphysics. It is a lot like Skepticism, but rather than saying that we will never know about the thing, it says that we will always be mistaken. For Meta-Ethics, this is the stance that all of our moral judgments will be mistaken. We will always be in error, hence the name. This is true even if we were to randomly chose an answer. For the **Error Theorist**, we will ask questions about a subject, sincerely try to answer them, but will always get the wrong answer. There is

no right answer. Applied in other fields, it claims that all of our judgments in those fields will always be in error. They come to this conclusion from three assumptions, or starting points, which get them off the ground:

First, they need to have the standard-issue feature of a moral nihilistic stance, this is that morality is not a feature of the world. In other words, there are no moral facts. Some could say that morality is a useful fiction or say that morality is self-contradictory or some other means. The second feature is a bit more precise to this stance. This is the claim that no moral judgments will ever be correct/accurate. This follows, using some fairly simple reasoning, from the first feature. For our judgments to be accurate/correct, they must correctly/accurately depict something. However, because of the core premise to Moral Nihilism, there's nothing for the judgment to correctly depict, so it can never be accurate/correct. The third aspect follows suit, they claim that it's pretty obvious that people try to use moral judgments, we act on them, and we think about them, but those judgments will always be in error, hence the name. Since our judgments can never be correct, they will always be wrong. The error theorist is making a pretty big claim. They aren't just trying to bash on social policies and individual actions. They are going all out and claiming that morality is a fiction.

They are essentially the atheists of ethics. Atheists hold that there are no religious features in the world, the error theorist holds the same about ethical features. The atheist holds that, try as you might, all religious doctrines are mistaken because (for example) God does not exist. The error theorist holds that, try as you might, ethical judgments/beliefs are all mistaken because morality does not exist.

### **For Error Theory to be Correct**

All error theorists hold that the basic mistake in moral thinking is that it depends on some absolute or objective truth about

morality. This truth applies to all of us, regardless of who or what we are. These are what some refer to this as ‘objective morality’ and ‘categorical reasons’.

To get error theory off of the ground, they need to convince us of 2 things. First, morality really does depend on these absolute standards (they would need to come up with the correct moral theory, examples are in Module 8). Second, supposing that it does, they have to show us that at least one of these is false, which would mean that the correct theory to describe morality is fundamentally and irreparably flawed. So, for the Error Theorist to prove themselves, they need to solve all problems in Ethics and then show that the stance which did it couldn’t be right. This is quasi-contradictory.

## Expressivism

Expressivism could, in principle, be a stance in other areas of philosophy and maybe, for some of them, it could make a solid stance, but I have not encountered it so much in my time and, when I have, it has almost exclusively been in Ethics. As I mentioned before, given the core flaws in trying to prove Error Theory (in the case of Ethics, in other fields this issue is not present), **Expressivism** is the stronger of the two. As it is a Nihilistic stance, it does come with two assumptions, but it differs from Error Theory in the last aspect.

First, as with the Error Theorist, the Expressivist claims that morality is not a feature of the world. But, it doesn’t say that it’s a useful fiction or anything like that. It also claims, but the same reasoning, that moral judgments are never accurate. But, it does not say that they are always wrong. I know that this last bit seems odd, so let me explain: The Moral Objectivist (Realist), Relativist, and Error Theorist all share one thing in common: When we make moral claims, we are, at least, trying to describe the world, pick out some feature in it. The Expressivist is different: They claim that when we make moral claims our intent is to express something,

not describe. They think that our ethical statements/judgments aren't the sort of things which can be true/false, correct/incorrect, or accurate/inaccurate. Take these two examples:

|                 |    |  |                 |                           |
|-----------------|----|--|-----------------|---------------------------|
| Murder<br>wrong | is |  | Blue is a color | Although it may look like |
|-----------------|----|--|-----------------|---------------------------|

(A) and (B) are stating the same sort of thing, and, in fact, were we to do a sentence/semantic diagram for these examples, they would look freakishly similar, according to the Expressivist, these are fundamentally different sorts of sentences. In linguistics, profanity is often claimed to not add any meaning to a sentence, but rather is an emotional or rhetorical expression; the Expressivist says, basically, that moral claims, like (A), are the same sort of thing, they are only emotionally/rhetorically meaningful. In the case of (B), it is expressing something true about the world, there's a collection of things, colors, and it's putting 'blue' in that collection (by one theory) or there's a thing in the world, 'blue', and assigning a feature to 'is a color' (by another theory, this is what I really work in). But, in the case of (A), we aren't assigning a feature to murder, according to the Expressivist. Rather it is like we are saying something like:

BOO murder!

Don't murder!

Let's not murder!

Wouldn't it be nice if we didn't murder?

Those sorts of statements aren't the kinds of things which are true or false, they are more like actions or commands or questions (for interesting content on statements as actions, check out J.L. Austin's *How To Do Things With Words*). If you have ever been to a highly emotionally charged (peaceful) protest, people screaming statements which boil down to "this is wrong" will feel more like "boo this thing" rather than something based on an intellectual process, which is evidence in support of this sort of stance.

### Three Concerns for Expressivism:

There are three major worries for Expressivism. Although, I will say, Expressivism is the most popular, it is by far the weaker of the two stances and it will need some time in the gym, so to speak, to really hold its own. The first worry for the Expressivist concerns how we really reason about morality. Take the following arguments:

All cases of hurting people are immoral.  
Torture hurts people.  
Therefore, torture is immoral.

All men are mortal.  
Socrates is a man.  
Therefore, Socrates is mortal.

In fact, we use the exact same style of reasoning in both cases, agree or disagree with it, the logic is exactly the same. But, according to the Expressivist, the first argument looks like this:

Hurting people - YUCKY  
Torture hurts people.  
Therefore, BOO torture.

We reason about morality logically, not emotionally (sometimes we do both, but when we are thinking clearly, emotions play a supporting role, they aren't the main character). According to Expressivism, however, we wouldn't do this, it would only be emotional. This observation, if correct, means that Expressivism can't be accurate.

The second concern for Expressivism comes from Amoralism. An amoralist is a person who sincerely makes moral claims but is completely unmoved into action by them. These people might be out there, this would be like Data from Star Trek, in that they would not feel the emotions associated with ethical

claims. If they do exist, this is a serious problem for the Expressivist. Expressivism claims that moral claims are emotional expressions. But, emotional expressions, most of the time, move us to action (even if we don't actually act, we still kind of want to). How can a person sincerely make a moral claim, which is supposed to be an emotional expression, without being moved to action? Basically, if a person can sincerely make a moral claim without being moved by it, then Expressivism is incorrect.

The final concern for the Expressivist concerns how we actually make these moral judgments, much like the first. The absolutist (objectivist/realist), the relativist, and the error theorist disagree on just about everything, but they do agree on one thing, that moral claims are, at least, trying to describe the world. The Expressivist denies this. They have to paraphrase the moral claims to not have them attempt to describe the world. But, if this were really the case, why wouldn't we just express our emotions and not conceal them? There are many statements, which are ethical in nature, which the Expressivist can't handle, many of which I have heard people make in the real world, for example:

I'm not sure whether torture is moral, but I think people smarter than me would know.

There's a difference between being moral, being required, and being praise worthy.

How much we punish should match how bad the crime was.

If war is immoral, then generals are not as good as they seem.

If we read these like the other stances would, then we have no problems, all are clearly understandable, but if we read them as the Expressivist thinks they should be, we are up a creek. We could all be lying or mistaken, but when we think about what we are trying to do when we make these statements, we



see that we are really trying to describe the world.

## MODULE VIII

*How Should I Act?*

## *Part 17: Normative*

### *Ethics*

Our focus in this module will be normative ethics. It is worth noting that many people confuse normative ethics and descriptive ethics. Descriptive Ethics is the study of how people do, in fact, think about ethics. Different cultures and people think about ethics differently, this is fine and totally normal. Normative Ethics is concerned how we should be motivated and how we should act. Confusing these two has lead many people to moral relativism, but it's rooted in a confusion, not a fact. Confusing these two things also is the main support for the Cultural Differences Argument which we saw before. The project in Normative Ethics is to find a theory which best explains our moral intuitions and best describes universally held moral standards. One way to think of the difference is that Descriptive Ethics is the realm of sociology, while Normative Ethics is the realm of philosophy. Different theories have been made to best describe our moral intuitions and are still being made, each improving on the last (so there is progress).

## Ethical Theories

As we saw in the last module, Moral Relativism just can't be correct. The reasons for supporting it contradict each other or it leads to truly awful beliefs. If we still want to be Moral Realists, then we are going to need to go with something objective, a theory of morality which has a set of rules or guiding principles to tell us what to do. Trying to answer this question has led down two different paths, the first we will cover belongs to the **Consequentialists**, who hold that we can tell whether an action was morally right or wrong based on the consequences, hence the name. In general, this branch thinks in terms of well-being and happiness/sadness. The second path people have taken in trying to get to an Ethical Theory goes belongs to the **Non-consequentialists**, who hold that the consequences of an action have no bearing on the morality. In general, this branch points to respect, free-will, the human spirit, and rationality to get the morally right actions. Sometimes you will find middle-ground theories, but these tend to come from the Consequentialists, not the Non-consequentialists. Both of these branches share one thing in common, though, they are trying to answer the question "what should I do?" There are other ethical questions out there concerning different questions, for example, the field of Virtue Ethics, found in Aristotle and most of Eastern Philosophy, is trying to answer the question "who should I be?", which is very different. It seems clear that a good person can make a moral error and a bad person can do the right thing. For this class, we are mostly only going to be concerned with the 'what should I do?' question, but other philosophy classes are out there which cover the other question.

For both the consequentialist and non-consequentialist branches of the tree, we can add further stances by asking various questions. For example, if we ask the question "Do I have moral duties to other people?" and get either "no" or "sometimes", then, likely, we are dealing with a flavor of ethical egoism, which is the stance that something is moral if it benefits

you selfishly. You might have heard of this stance before, but by a different name. The most famous supporter of this was Ayn Rand.

## The Scientific Method for Ethics

When you are given a general Ethical Theory, there is a scientific style method for testing it. When philosophers propose a hypothesis about what makes an action right or wrong (good or bad) (in this context, we call them theories), they come from looking at the facts about people's general moral intuitions and why we think a person did right or wrong in their actions/why a person is a good or bad one. The 'why' aspect is key. Most of the time, our moral intuitions are a gut instinct, from a sort of mental-shortcut. The vast majority of the time, moral disagreements can be settled by exposing the shortcuts. This moves the moral debate into a metaphysical one, one which we can debate about.

The philosopher then looks at the predictions made by the theory, what it tells us to do in some case, and then checks to see whether it lines up with the world around us, and the generally accepted moral duties which we have. This is not just looking at the philosopher's own culture, rather it also comes from looking at the perspectives of other cultures, for example, there have been very interesting developments from comparing the moral intuitions of the Ashanti People of Ghana and how they line up with general moral principles, and if the prediction and the intuitions don't line up, the philosopher amends the theory and tries again. This is how we do things in the real world; we are told some principle, we follow it until it makes a wrong prediction, then we adjust and amend, and repeat. For this class, we will be using this argument structure:

If theory A is correct, then action B is what I should do.  
Action B is not what I should do.  
Therefore, theory A is not correct.

This also comes in a slightly different form. We chose which according to the prediction of the theory.

If theory A is correct, then I shouldn't do action B.

I should do action B.

Therefore, theory A is not correct.

We start with what the theory claims would be the case and then see whether it lines up with the data (for this class, the data is your gut instinct when it comes to the ethical case. There are areas in psychology and philosophy which actually test people's moral intuitions by, for example, putting them in a VR situation and measuring the response). This system works for most of the theories covered in this class. The trick is to figure out what the theory says about a particular case and then see whether you think that this is correct.

Sometimes, however, it is harder to pinpoint something that the theory gets wrong in terms of actions. In these cases, it pays to look into how the theory chooses whether an action is right or wrong. Those cases don't tend to involve me making a fun story, which I do for almost all of them (there are exceptions, Divine Command Theory, Virtue Ethics, and Feminist Ethics). The more applicable the theory is to individual choices, the easier making a case where it makes a prediction is.

## *Part 18: Utilitarianism*

Utilitarianism is based on the idea that happiness is good. It is Consequentialist in nature. Utilitarian thinkers have traditionally understood happiness in terms of pleasure and the absence of pain. The general ideas behind this theory are found the world over, for example, you can find this theory being discussed and spread in Ancient China, in the Mohist School, between 479 BCE and 221 BCE. It was first really formalized by Jeremy Bentham (for a fun fact about him, look up the Auto-Icon) and then was further developed by John Stuart Mill, who is its best known advocate. Mill, characterizes **utilitarianism** as the view that “an action is right in-so-far as it tends to produce pleasure and the absence of pain.” Mill, also, has a fascinating life story as well, if you care to look into it. If happiness, conceived of as pleasure and the absence of pain, is the one thing that has positive value and pain/sadness being pleasure’s opposite (the one thing with negative value), then this criterion of right action should follow fairly quickly. This leads us to the Principle of Utility.

In any given scenario, the actions we make in that scenario will have consequences. We assign those consequences value based on the amount of happiness (pleasure) caused and the amount of sadness (pain) caused, for all beings affected, that action’s 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 need 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.

There are a few things which are worth noting about Utilitarianism. First, this has no room for self-interest bias. You are not special in the moral figuring. To use this theory, you need to count your own wants and desires equally to all other beings who could be affected. For example, suppose that I really want a million dollars, so I rob a bank. If I only counted myself in the equation, then it would tell me that this is the right course of action (assuming I don't get caught). However, there are other people and I am not, as it were, an island. The grief and suffering caused to others by my action would greatly outweigh the benefit to myself, and that makes the action wrong.

Second, this theory only requires that you cause the most good (once the bad is subtracted). This good does not need to be evenly distributed. For example, suppose that a government has a surplus and the president wants to use that money to help the citizens. The Pres. has a couple of choices. She could distribute the money in the form of gas-cards to the top 90% of the population. This would cause some happiness for the majority. The other option is to use the money to help fight starvation, provide education/work training, and otherwise help the bottom 10%. While the first option causes good things for the most people, in the grand-scheme of things, dedicating the money to the bottom 10% is the better option because the suffering removed is greater than the pleasure which would have been added.



Third, this theory does not make a distinction between long term and short term consequences. This makes sense most of the time, but it could be an issue for the theory when it comes to practical use. For example, allowing your child to eat all of the candy and junk food they want will, likely, maximize their happiness in the short term. In the long-term, however, there are a bunch of different health problems and behavioral issues which arise from that sort of diet, making it wrong to feed a kid a ton of junk food.

Here is an example of a course of action I could take, supposing that I came into some money:

|                   |  | Hedons<br>(happy-points) | Dolors<br>(unhappy-points) | Total |
|-------------------|--|--------------------------|----------------------------|-------|
| Buying New Shoes  |  | 10                       | 500                        | -490  |
| Donating to Oxfam |  | 500                      | 10                         | 490   |

For this case, suppose that I have a choice between buying myself a new pair of shoes or donate to the charity Oxfam, which is there to alleviate world hunger and poverty. Think of 'hedons' and 'dolors' as made up units of pleasure and pain. I like to think of them relatively, so we can roughly say that donating to Oxfam will cause 50x as much good as buying the new shoes, and it will also cause a 50th of the suffering. According to the Utilitarian, this is the route you should go, make the most hedons and the least dolors. Remember, these are not actual units, they are just conceptual tools to help understand the theory. You can use different unit sizes for different cases, just keep them the same in the same case.

Utilitarianism has no room for the individual making the choice seeing themselves as special, weighing their happiness as greater. It's the total happiness that matters, not just your happiness. So utilitarianism can call for great personal sacrifice. The happiness of a child during their lifetime might require great personal sacrifice from the parent/caretaker during the

first few decades. Utilitarianism says that all beings are equal, morally speaking, so long as they are equal in their ability to feel pleasure and experience pain.

Likewise, Utilitarianism has no room for favoring the immediate consequences over the long term. In this primitive form, an actions utility covers all time after the action. So, while it might maximize a small child’s pleasure in the short run to be given ice cream whenever they want 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.

Here is another example of this at work. In this case, each of these choices on how to spend my money have different ripples through time and different amounts of people that they affect. For an interesting exercise, think about how long the results of the actions will help others and how many they help.

|                              | Hedons<br>(happy-points) | Dolors<br>(unhappy-points) | Total |
|------------------------------|--------------------------|----------------------------|-------|
| Buying a new laptop          | 40                       | 20                         | 20    |
| Donating to a foodbank       | 80                       | 20                         | 60    |
| Give money to broke relative | 60                       | 20                         | 40    |

Like I said, you can use different sizes, just keep them the same in the same case. As we can see, the amount of suffering caused by my action is around the same across the board. But the amount of pleasure is different, the one with the most is to donate the money to a food bank. According to the utilitarian, this is what I should do.

## Why People Like This Theory

There are several reasons why people tend to like this theory. In other words, there are several positive aspects of it which

seem to get morality right. The first reason is that the theory is impartial. It does not have in it any sort of bias for or against a person or even animal. As Jeremy Bentham put it, it doesn't matter whether they can speak or whether they can problem-solve, rather it only matters whether they can suffer. As a result, it doesn't matter whether you are rich, poor, black, white, male, female, man woman, part of some religious sect, or none at all. Morally speaking, you are equally worthy of consideration. Sometimes we take this sort of equality for granted, but when this theory was being formulated, it was not the norm, and even today, equality of this sort is fought for. You don't need to argue for it in utilitarianism, it comes built in.

The second reason people like this theory so much is it gives ample justification for moral claims. Utilitarians are no strangers to being controversial. Classical Utilitarians have been on the front lines of many periods of history: the abolishing of slavery (Bentham, 1748-1832), equal rights for women (John Stuart Mill, 1806-1873), and animal rights (Peter Singer 1946-present) to name a few. That said, utilitarians can give reasons why many of our deeply held moral beliefs are correct. They see this as a major plus for the view. For example, many of the things which we find repugnant morally, slavery, killing innocent people for no reason, and others, all tend to do more harm than good. At the same time, things we have strong moral feelings in favor of, helping others, telling the truth, bravery, and others, all tend to make the best outcomes.

Third, a really important feature in a theory is that it can provide practical advice about what to do in a given situation and how to resolve conflicts. When you are in a situation where you are not sure what to do, the utilitarian can apply their theory and tell you what the best option is, there's something in the world which they can point to, the outcomes, and say which one is the best. Similarly, when two groups are in disagreement about what the morally right thing is, the Utilitarian can point to certain base-level ground-work facts about the case and re-

solve the issue.

The fourth plus which this theory has going for it is its flexibility. As I mentioned in the part about justification, certain rules tend to produce the best outcomes, so the utilitarian supports them. But, those rules are not absolutes, they are there because they tend to get the best result. Sometimes, we need to break those rules in extreme situations. Utilitarians are just fine with this and can easily give the reasons why those rules should be broken. Its flexibility lets it work in most any situation. Take this historical case as an example: In the winter between 1846-1847, members of the Donner Party, traveling west, found themselves in heavy mountain snows. About half of the 87 members of the party died after food and supplies ran out. Those left had a terrible choice: Starve to Death or Eat the Remains of their fellows. Some may think that the rules against cannibalism are absolutes. Not to be violated under any conditions. The Utilitarian disagrees. This disagreement is not based on cannibalism, but rather that no act-type is absolutely wrong. While it's wonderful that many of us are against these sorts of actions, utilitarians understand that desperate times call for desperate measures.

## Concerns for Utilitarianism

Now, this theory is really good at a lot of things, it is especially useful in high-stakes situations, but there are some concerns which need to be addressed and areas of improvement. The first is an epistemological worry which should have caught your eye in the last page. Often, We don't know (and sometimes we can't know) what the consequences of our actions will be in the long term. Sometimes, even in the short term, we may be uncertain about what will happen. For example, what if one of the people affected is particularly strange and gets happiness from pain? These worries point to a concern about the practicality of the theory. It's difficult to use it in our daily lives.

This worry/concern does NOT take Utilitarianism out of running for being a good Normative Ethical Theory, though. 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. But they aren't the only worries for this primitive form of Utilitarianism (Act-Utilitarianism). There are other worries which are troubling for the very core of the theory, the Principle of Utility.

### **The Porky the Pig Problem Problem**

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 the Utilitarianism to be unfit for humans because it recognizes no higher purpose to life than the mere pursuit of pleasure. This objection only applies to treating pleasure and pain in their most basic, animalistic, senses. This was not originally mine, but a version of it was given as a two day lecture when I was in community college. The very original version is not as fun to give and it was published by G.E. Moore in 1903. If you want to see one other objection, one which you could write about in the assignment as well as the original objection to hedonism (which is a more primitive form of utilitarianism, without special amendments, utilitarianism falls into both of these problems). But, here we go:

Way deep in the back woods, there lives a pig farmer named Porky. Porky raises prize winning swine and sells them to cover his basic needs. He has no wife/husband and doesn't really want one. One day, when he was particularly bored, he noticed his swine having fun wallowing in the mud. Thinking that this looked like a good time, he stripped down and jumped in, having a whale of a time. As time goes on, Porky needs more entertainment than merely wallowing with his sows. He starts thinking "man, that's a pretty little piggy". Over time, late in the evening, his neighbors start hearing strange sounds coming from the direction of Porky's shack. They think nothing of it, maybe Porky is just doing some late night breeding to get better piggies for market. Porky is doing some late night breeding of a sort. He is engaged in bestiality! And O! Is he enjoying himself!

With all this information about how awesome and pleasurable Porky's life is with his piggy time, we now have to ask "is what he is doing moral?" Well, according to Utilitarianism, if we take pleasure to be in the animalistic sense, it is.

The morally right action is the one which produces the greatest amount of happiness/pleasure for the greatest number of people.

If Porky doesn't engage with his pigs in this way, he will have very few pleasures.

If Porky does engage with his pigs in this way, he has a lot of pleasures.

If he has a lot of pleasures, then the greatest amount of happiness over the greatest number will be served.

Therefore, the morally right action is to engage with his pigs in this way.

But this can't be right. The core of it is that there is no consent (among other things), which means that it can't be right.

If Utilitarianism is correct, Porky's actions would be morally permissible.

Porky's actions are not morally permissible.

Therefore, Utilitarianism is not correct.

### **The Reply to Porky**

Because of this objection, Mill and others don't take pleasure in the animalistic sense, but rather in a far more broad sense, where the other intellectual and emotional pleasures are taken into account. Mill responds that it is the person who raises this objection that 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. This response seems OK to some, but others argue that a sufficient amount of physical pleasures can, in principle, outweigh the intellectual.

### **The Utility Monster Objection**

One objection to Utilitarianism isn't concerned with what it measures to determine morality or even how it is measured, rather this objection concerned how it determines the morally right action given the outcomes. As Act-Utilitarianism sits right now, it claims that the morally right action is the one with the highest outcome given the available options. This, however, leads to the possibility of a utility monster. A utility monster is a being which receives a massive quantity of pleasure (happiness, the good) from consuming resources, higher than any other being, by a significant margin, often at the expense of others. Just a cursory glance over the numbers would have Act-Utilitarianism claim that such a utility monster is doing the right thing in exploiting or harming others for their

own benefit, because of the massive amount of good which they receive. To put this idea as an example, take the following case:

Peter Singer<sup>a</sup>, late in the evening, while he is resting at home, hears loud banging and commotion coming from his basement. So, he grabs his flashlight and goes down to investigate. He sees a green, round, being with arms and legs tearing apart Singer's plumbing causing massive flooding. "What are you?" Singer exclaims, "and what are you doing to my pipes?" The creature pauses their destruction for a moment and turns to Singer. "I am destroying your plumbing" they explain, "you see, I love wrecking pipes, far more than any suffering caused to you. I am a utility monster."

---

<sup>a</sup>Peter Singer is an Australian philosopher and is best known for his, seemingly, extreme views regarding Utilitarianism. He holds that the theory is correct and applies it to many contemporary issues. For example, one of his foundations, The Life You Can Save, tracks the spending habits of various charities and connects donors with the one which will get the most 'bang for their buck' in the issue which they are concerned with. Singer, holding true to this belief, lives well below his means and donates substantial amounts to charity. He is mostly concerned with world hunger and poverty, but he has been known to be outspoken about animal 'rights' and welfare.

In this case, most of us would say that it's wrong for the utility monster to destroy Singer's pipes. There are concerns about personal property and there are concerns about the harm done to Singer. However, Act-Utilitarianism, without any modifications, measures pleasure and pain with the same metric. One unit of happiness cancels out one unit of sadness. As a result, so long as the Utility Monster didn't have any other options which would cause a higher total, Act-Utilitarianism would say that they did the right thing. This does not square with general intuitions about morality.

Act-Utilitarianism gets the wrong result in this sort of case.



And this objection is so basic, to the core of Act-Utilitarianism, that it might lead us to a change in theory, amend the theory to better fit our intuitions. One possible alteration is to change what is measured to determine morality. In this case, rather than measuring happiness and sadness equally, you only measure the sadness caused by the action or inaction. This is called negative utilitarianism. Negative Utilitarianism was proposed, most notably, by Karl Popper in his work *On the Open Society and Its Enemies*. Popper states that the morally right action is the one which minimizes suffering, rather than maximizing pleasure. Going this route avoids both the Utility Monster Objection as well as the Porky the Pig Problem but it also leads to certain other problems, if you take the letter, rather than the spirit, of the moral theory.

### **The Organ Harvest Problem**

This problem is often seen as a more gruesome version of the Trolley Problem for Ethics. This objection to Act-Utilitarianism stems from the idea that only the results matter, the ends justify the means. I personally have other versions of problems like this which involve framing an innocent person, forging evidence, and rigging elections, all of which have (due to the situation that they are in) the best consequences. Consider the following case:

Bob goes to the doctor for a check up. 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.

The question for the Act-Utilitarian is "what should the doctor do?" According to the theory, it seems that the good doctor should quickly and as painlessly as possible kill Bob and harvest his organs, getting them to the 6 other patients as quickly as possible. This is because, to quote Spock, "the needs of the many outweigh the needs of the few." The overall outcome of letting Bob go is the value of one average good life minus the values of six average good lives and the overall outcome of killing Bob is the value of six average good lives minus the value of one. But this can't be right. Our intuitions clearly speak to the immorality of this. Act-Utilitarianism, again, gets the wrong result in this sort of case.

## Rule Utilitarianism

After having the concerns with Act-Utilitarianism, more particularly, cases similar to the Organ Harvest Problem, many people have thought to amend Utilitarianism to better fit those cases. These amendments are perfectly fine, but they have reached the point where they have developed into two different theories in this family. The first, which we will cover in this class, is called Rule Utilitarianism and the second, which we will not be covering, is called Negative Utilitarianism.

Rule Utilitarians move away from the idea that we need to max-out the utility of a given action. These Consequentialists look at the rule which was followed in making the action (hence the name). They say to act according to the rule which, if followed by everyone, would lead to the greatest utility generally. So, for example, if we look at the Organ Harvest Problem, a rule which had doctors kill their patients to save others (kill one to save six) would not have a very high utility if all doctors followed it. People would stop going to the doctor, they wouldn't trust medicine, etc. On the other hand, a rule which had doctors do no harm (forcing a difference between doing and allowing harm) would lead to a greater utility overall. We trust our doctors, in part, because of this. So, a move to this sort of Rule Utilitarianism might be a step in the right direction. But, there is a third rule with an even bigger utility.

Consider the following rule: 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 Rule Utilitarian must find a principled way to exclude certain sorts of utility maximizing rules. Rather than pursuing this further for the Utilitarian, I want to consider further just how simple Act Utilitarianism goes wrong in Bob's case (and I say this as a person who honestly thinks Consequentialism is correct). Utilitarianism evaluates the goodness of actions in terms of their consequences. 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.

## *Part 19: Kantianism*

Like Utilitarianism, Immanuel Kant's moral theory states that there is an absolute moral rule. But, unlike Utilitarianism, it states that the consequences of the action don't matter morally. It is, however, grounded in a theory of intrinsic value. Rather than going with the Principle of Utility, Kant's theory has it that the only thing with moral worth is the Good Will, which we find in persons. Persons, for this theory are autonomous rational moral agents. This theory, from the very start, makes a certain metaphysical assumption: People have free will. This can't be the kind of free will proposed by the Compatibilist in the sense we have seen in this class, as Kant was fundamentally opposed to it, so the sense of free will must be the Libertarian sense. There have been attempts to make a Kantian Compatibilism, but those seem to have been unsuccessful.

**Kant's Moral Theory** rests on this notion of the **Good Will**, so we should be clear about what that is. 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 eloquent sentence places this 'Good Will' at the center of Kant's value theory. The one thing that has intrinsic value, for Kant, is the autonomous good will of a person. But we can't understand Kant's Good Will in the ordinary sense. In everyday discourse

we might speak of someone being a person of good will if they want to do good things. But this would be a Consequentialist take. On Kant's view, the person with Good Will wants good things out of a sense of moral duty, not just some habit or tendency.

The naturally generous philanthropist doesn't demonstrate their good will through their giving according to Kant, but the selfish greedy person does 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, to be worthy of dignity and moral regard, for Kant, we need to have the ability to see what our moral duty is and act according to it. For Kant, then, in order to be worthy of moral consideration, we need to be able to act in a way which is opposed to our desires/conditioning. 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's version of the Principle of Utility, his fundamental principle to guide our actions and thereby give people the dignity and respect worthy of their Good Will is called The Categorical Imperative. An imperative is a command or order given. Kant, and others, explain this kind of imperative by contrasting it with another kind, a hypothetical imperative. A hypothetical imperative is a command, but it only applies conditionally according to your desires, what your goal is. For example, "if you want to avoid traffic, leave 15min early." This imperative tells you what to do if you want to avoid traffic, but it fails to tell you what to do in a case where you don't want to avoid traffic. A categorical imperative (though Kant thinks there's only one) is different in that it tells you what to do regardless of your goal or desires. It applies according to the kind of action you are

taking, not why you are taking it. Kant also holds that if there is a moral law, it will be the Categorical Imperative. Treating it that way, moral reasons must always overshadow any other sort of reasons which can be given. 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.

Although Kant gives 4 different statements of the Categorical Imperative and claims that they all boil down to the same basic idea, there is some debate about whether they do. Two of the formulations just seem to be restatements of the other two, and the debate is over whether these two can be formed together. Once you see them, you will see that they certainly don't look like they are expressing the same idea. The first formulation which we will cover is called The Principle from Humanity, for this class, though other names for it are out there. The second formulation is called The Principle from Universalizability for this class, though, again, other names are out there.

## The Principle From Humanity

Always treat persons (including yourself) as ends in and of themselves and never as a mere-means

This formulation tells us to treat individuals as ends in themselves, but what does that mean? How do I use people as mere-means? How could I use myself that way? This formulation or principle is noted for really highlighting the notion that people are intrinsically valuable. To say that persons have intrinsic value is to say that they have value independent of their usefulness for this or that purpose. They are not a tool or resource

which you can use without consideration of their worth. the Principle from Humanity does not say that you can never use a person for your own purposes (using them as a means). If this were the case, you taking a class from me would be morally wrong. It tells us never to use others as a mere-means.

## Means Vs Mere-Means

We treat people as a means to our own ends in ways that are not morally problematic all the time. When I go to a grocery store to pick up some food, I treat the clerk as a means to my end of buying food. But I do not treat that person as a mere-means to my own end. I accomplish my end of buying food 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 someone as a mere-means to our own ends if we force them to do our will, or if we deceive them into doing our will. We are, in a sense, using them without their consent, where it would not be possible for them to consent. Sometimes we use people as mere-means when we don't take their goals into account in our interactions with them.

Suppose that I have the goal of building a tower with my name emblazoned on the top. I have the money to do this and I hire workers to build it (they have the understanding that they will get paid when the job is done). Just before they finish the work and would expect to get paid, I take all of my money, put it in an off-shore account under my son's name and declare bankruptcy. Using the laws in this regard, I get the contracts waved so that I don't pay the workers. Later, when the dust settles, I transfer the money back into my name.

In doing this, I have used those workers as mere-means. I



have the goal of getting the building, they have the goal of getting paid. In my interactions with them, according to Kantianism, I should have, in a sense, made their goals my own and had the intermediate end of paying them for the work. In the case above, the workers and I do not share the same end, so I am using them merely as a means.

Coercion and deception are paradigm violations of the categorical imperative. In coercing or deceiving another person, we disrupt their autonomy and violate their will. This is what the categorical imperative forbids. Respecting persons requires refraining from violating their autonomy.

Here is an example of this sort of theory at work, where it seems to get the right answer:

To take a person's life, liberty, or legitimately acquired property without that person's consent is to use that person as a means to an end (if they give consent, then they are being treated as an end). It is to treat a person as a tool or resource placed here for your convenience.

It is always wrong to treat a person in that way.

Therefore, each person has a moral right to life, liberty, and property, regardless of the consequences.

So, killing is always wrong, so is stealing. Another example of this formulation getting things right comes up when we talk about slavery (though the Utilitarians were on the front-lines against slavery before the Kantians):

Each person is intrinsically valuable regardless of race, religion, ethnicity, or national origin and deserves to be treated as such.

Slavery does not treat people as an intrinsically valuable being.

Therefore, slavery is morally wrong.

## Problems for Humanity

### Taxation

Taxation is the taking of a person's property and using it to benefit others, without their consent.

Taking a person's property without their consent is treating them as a mere means to an end.

Treating a person as a mere means to an end is always morally wrong.

Therefore, taxation is always morally wrong.

Some of my more conservative students may like the sound of that, "Taxation is Theft" they tend to shout. But many of you will have a problem with the consequences of taxation always being morally wrong. Public schools, as they are paid by taxes, gone. Most public roads, if they do not lead to some person's business and they did not pay for it, gone. The cost of your college education will sky-rocket, as they are subsidized by taxes. And many others.

### Lying

This is an interesting case, and one which Kant himself thought was right, as in he thought that the theory got the correct answer here, however, this does not line up with most people's moral intuitions about the case. Take this case as an example (which will appear in the discussion for this module):

Your roommate is in the shower after having a late night with a recently gained boy/girlfriend, who had gone through a recent bad break-up (their former significant other is "the ex", this is not the ex of your roommate, for clarity) . You hear a knock on the door and go to answer it. You find the ex standing at the door with an axe, murder in his/her eyes. The ex asks you "Where is [insert roommate name]?" Neither of you can hear the shower running. You know that if you tell the ex, then they will rush right through you/knock you out/whatever and get to the roommate killing them.

According to The Humanity Principle, we should never use others as a mere-means. But, does lying count as a mere-means? According to Kant, it does. Kant says that in lying, you are misinforming one person for the benefit of another (or yourself), without appreciating their intrinsic worth. This is always wrong, according to the principle, so lying, no matter what, is always wrong. Sure, you could refuse to answer, slam the door, and so on, but you can never lie. This means that lying to the axe-murderer is wrong. So, if you need to speak, then you must tell them that your roommate is in the shower, and more than likely, deal with the Psycho scene later.

However, this is an issue, because it certainly seem right that there are cases where lying is permissible, so this seems wrong, too hard lined.

## The Rendering Aid Problem

This particular problem addresses something interesting in this formulation of the Categorical Imperative (and this feature should be found in any rephrasing of it, as it's core). In it, it states that we should not use ourselves as mere-means. Since Kant explicitly stated that we shouldn't, it must thereby be possible to use ourselves as mere-means. This was quite the

puzzle for me, personally, how could I possibly do something without taking my own goals into account? How could I force myself to do something against my will? The examples which Kant himself gives have not aged well, they involve sexual acts and suicide, which I don't want to use for this course.

I have hence asked around and some examples do seem problematic and I will look at the issue of rendering aid.<sup>2</sup> Often we glorify people who help others at the cost of their own wants and goals. In such a case, they are not taking their own goals into account and using themselves to benefit others. This, by its very nature, would be the person using themselves as a mere-means. Here is an example, in an ordinary case:

Suppose that I have the goal of buying a new video game, The Elder Scrolls 20 or some such, and I know that given the sheer demand for the game, if I am not there just as the shop opens, it will be sold out. As I am driving to the shop early, I notice a young man having a hard time changing a tire<sup>3</sup>. Feeling as though I should help, I pull over to render aid. This prevents me from making it to the shop on time and violates my end.

In this case, it would seem that I have used myself as a mere-means and cases of rendering aid like this would be morally wrong.

## The Principle from Universalizability

Act only on the maxim that you can consistently will to be a universal law

This version is also known as the formula of the universal law. The maxim of our action is the base-level reason or principle that determines what we are doing. We act for our own reasons and different goals might lead to similar actions. For example, a person might wash their clothes regularly because they

don't want to smell bad while another person might do the same because they don't want their significant other to complain about the stack of clothes near the closet. Though they have the same action attached to them, the maxim behind the action will be different. For Kant, intentions matter and this formulation really gets at this point. 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. Whether an action is right or wrong is determined by the actor's intentions or reasons for acting.

It should be noted that a **maxim** should not include your personal wants and desires, those would make it a hypothetical imperative. A maxim should be written/phrases as something along the lines of "I will A in order to realize C." Where A is the action you want to perform and C is the reasoning behind the action. So, suppose that I am deep in debt to a loan shark because of my gambling habit. I go to the bank to get a loan to pay the loan shark and save my knee-caps. The maxim for my action here would be something along the lines of "I will lie in order to get money."

According to this formulation, 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. Basically, ask yourself "am I making a special exception for myself?" "could anyone in my situation do this?" If anyone with similar desires could do what you are doing and accomplish them, then you are morally in the clear, otherwise you are morally up a creek.

Here is an example of this thought at work:

Suppose that I really want to win at a game, so I think about cheating. The principle I go on is “whenever I want to win, I will cheat in order to do so”. But, if everyone did this, the concept of a game would be mute, no one would play the game by the rules, so cheating would not be cheating. This is a contradiction, so cheating is always wrong.

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. Problems for this one

## The Traffic Jam Problem

Suppose that I regularly get caught in traffic at 6:45AM. I know that if I leave 15 minutes earlier, I will arrive where there's traffic at 6:45AM at 6:30AM, missing the traffic. So, I ponder the following:

Whenever I want to avoid traffic, I will leave 15min earlier.

But, everyone wants to avoid traffic, so what would happen if everyone did this? If everyone wanted to avoid traffic and left 15min earlier, it is reasonable to say that the traffic would not be at 6:45AM, but now at 6:30AM. This means that if everyone left early to avoid traffic, they would not avoid traffic. This is a contradiction. If I leave early, I avoid traffic and if everyone leaves early, no one avoids traffic. Therefore, according to Kant, me leaving early is morally wrong.

## The Breaking Promises Problem

Suppose that I promise my mother on her deathbed to sing and play a certain song on my banjo at her funeral. She asks me to play/sing *In Hell I'll Be In Good Company*, by The Dead South. But, on the day of the funeral, I can't bring myself to play such an inappropriate song, the lyrics in this context would make everyone's grieving worse. So, I leave the banjo to the side. The principle you are going with is "whenever I make a promise that I don't want to keep, I will break that promise." Now, what would happen if everyone broke promises they didn't want to keep? If everyone broke promises they did not want to keep, then the very notion of promising would go out the window. If there are no promises, you can't break them. That is a contradiction. If you promise, you break it. If everyone broke promises, there would be no promises. If there are no promises, you can't break them. Therefore, if everyone broke promises, they can't break promises. So, breaking promises is always morally wrong.

But in not playing that song, did I really do something wrong? Most will say no.

# *The Moral Status of Bloodbending by Davis Smith*

## Introduction

During their travels around the world of Avatar, our heroes have encountered many people with strange or unique bending abilities. For example, they encountered Combustion Man, who could make things explode with his mind.<sup>120</sup> None, however, are more troubling and ethically interesting than Hama, a waterbender from the Southern Tribe who can, on a full moon night, bend the very blood in a person's veins and manipulate their body to do her bidding.<sup>121</sup> Such a violation of a person's control over themselves would, rightfully, be the stuff of nightmares. However, are all cases of this bloodbending wrong? To answer this question, we turn to two well-established theories in

---

<sup>120</sup>First seen in "The Headband," (23:24-23:48) but his abilities are displayed for the first time in "The Beach" (15:21-16:16).

<sup>121</sup>"The Puppetmaster," creator Michael Dante DiMartino, author Tim Hedrick, director Joaquim Dos Santos, Nickelodeon Animation Studio, 2007.



Ethics, an area of Philosophy which, among other things, tries to answer the question “what is the moral thing to do?” Almost as if by design, bloodbending works as a fantastic example of how these theories determine the morality of an action and how they can come to radically different conclusions on a single case.

The episode gives us four different situations in which bloodbending is used and each can serve as examples for the different ways one can evaluate an action to determine its morality. In all four cases, the bender is forcing a person to do something against their will. So, if our moral intuition or the theory gives different responses to the question “was that the right thing to do?”, then it is not the manipulation of a person which makes that difference. Because the cases are so different, if an ethical theory gives the same response to all four cases, then the manipulation is the cause.

## The Ethical Theories

To start us off, I would like to give a basic overview of the different ways in which one could classify ethical theories, theories about whether some action is right or wrong. First, we can classify them according to what they use to make their determination. For this, there are two natural categories. On one side, we have theories which make their moral evaluation according to the consequences, the results of the action. These are called Consequentialist theories. The ultimate command for a Consequentialist theory is to leave the world a better place. Theories in this family differ in how they determine ‘better’ and in how they measure it, but they all share that primary command. On the other side, we have Non-Consequentialist theories. As the name implies, these theories are the opposite of the Consequentialist ones. They hold that the consequences are irrelevant to the morality of an action. Some actions, according to these theories, are just wrong, regardless of the outcome.

We could also classify them according to whether they are

concerned with individual instances of an action or kinds of actions. Act-Type theories start by asking about the kind of action it is. These theories tend to move from general principles or commands about categories of actions and use those to determine moral status of an individual case. For example, one could have the general principle “lying is always wrong” or the command “don’t lie” and from that determine whether some case of lying is wrong (it is, according to that principle). On the other hand, some theorists think that this approach is too impersonal and not down-to-Earth. This is where we get the Act-Token theories. Rather than moving from some general principle to make a judgement, Act-Token theories make their evaluations by just looking at the case in isolation. These theories have no problem claiming that most of the time lying is wrong, but they are more than willing to make exceptions and give reasons for those exceptions.

These two different ways of classifying ethical theories gives us four different, generic, types of theories. In this paper, we will be focusing in on the two major ones, Utilitarianism (a Consequentialist Act-Token theory) and Kantianism (a Non-Consequentialist Act-Type theory). Because of their radically different stances regarding the measurement of morality and what is to be measured, these theories serve as a wonderful foundation to explore the moral status of bloodbending.

## The Utilitarian Account

Jeremy Bentham (1748-1832),<sup>122</sup> John Stuart Mill (1806-1873),<sup>123</sup> and Harriot Taylor Mill (1807-1858)<sup>124</sup> built off each

---

<sup>122</sup>Jeremy Bentham, *An Introduction to the Principles Of Morals and Legislation* (The Online Library of Liberty, A Project Of Liberty Fund, 2011).

<sup>123</sup>John Stuart Mill, *Utilitarianism* (lccn.loc.gov/11015966: Parker, son, / Bourn, 1863).

<sup>124</sup>Harriot Taylor Mill’s husband was J.S. Mill and by his own words, she was heavily influential in his formulations and refinements of Utilitarianism.

other and formalized a theory of morality called Utilitarianism. This account of the permissibility of actions is Consequentialist in nature and, though it is possible for this account to use act-types, Mill and Bentham both certainly preferred focusing on act-tokens. These two features, almost immediately, give us a possible answer to the question “is bloodbending always wrong?”, namely, “no.” In addition to this, when we think about the more general question about whether it is permissible to violate a person’s autonomy, this theory would say “sometimes.” In fact, because it is an Act-Token theory, it will always allow for some exceptions to this kind of general rule.

There is one principle, one overarching metric, according to Utilitarianism, which cannot have exceptions, and which determines the morality of an action; The Principle of Utility.<sup>125</sup> This says that the action you should do is the one which results in the greatest amount of happiness (or well-being) once the suffering (sadness, pain) has been subtracted from it. It should be noted, however, that the doer’s interests and the effects on them are not treated as especially important. Your well-being is treated equally to all other human (and potentially non-human) souls. It is part of the core command of this theory that you think of others and ensure that, all else being equal, their lives are made better by your actions. Morally speaking, this theory has it that all beings are equal in so far as their ability to suffer or experience happiness is equal.<sup>126</sup>

The world of Avatar has no shortage of examples of Utilitarian style thinking, seemingly, getting morality correct, outside of cases of bloodbending. To start off, in Book 1, Episode 5, we first encounter one of my favorite side characters, the Cabbage Merchant. As he is attempting to enter the city of Omashu, we see the guards, without regard for the merchant’s desires or the effects of having a cabbage merchant in the city, use earthbending to launch his cart into a canyon.<sup>127</sup> Given what we know

---

<sup>125</sup> Bentham pg.12.

<sup>126</sup> Bentham pg.245.

<sup>127</sup> “The King of Omashu,” creator Michael Dante DiMartino, author

about the Cabbage Merchant and the results of this, we can say that the guards were wrong in launching the cart. In that same episode, we find Team Avatar preparing to use the delivery system of the city as a slide, for their amusement. In doing so, they cause a lot of damage to various people's property and they cause even further distress to the Cabbage Merchant.<sup>128</sup> This destruction and the suffering, for lack of a better word, caused certainly outweighs the enjoyment had by the team. It follows from this that Utilitarianism would say that what they did was wrong.

In Book 1, Episode 11, we encounter rival tribes (with radically different personalities) each seeking passage through a dangerous canyon.<sup>129</sup> The guide warns them that bringing food into the canyon will attract the carnivorous and dangerous animals. Both tribes end up bringing food into the canyon, believing that the other will. Given the risk that they are putting the people, collectively, in by doing so, we can see that this, like before, outweighs the benefit of not going a day (or so) without food. Utilitarianism, as a result, says that the tribes did something wrong here. To finish off our set of examples, we turn to Book 2, Episode 2, where Iroh is deciding about whether to make tea out of a strange plant.<sup>130</sup> As he put it, this plant is either "delectable tea or deadly poison." If we suppose that it is a 50-50 shot, we need to ask whether the temporary enjoyment from such a "heartbreaking" tea is equal to the suffering caused by his death. Since we know the negatives which would come from Iroh's death, we can say rather confidently, using Utilitarian reasoning, that Iroh making the tea was not worth

---

John O'Bryan, director Anthony Lioi, Nickelodeon Animation Studio, 2005. 02:53-03:07.

<sup>128</sup> *The King of Omashu* 04:54-07:29.

<sup>129</sup> "The Great Divide," creator Michael Dante DiMartino, author Aaron Ehasz, director Giancarlo Volpe, Nickelodeon Animation Studio, 2005. 00:00-23:37.

<sup>130</sup> "The Cave of Two Lovers," creator Michael Dante DiMartino, author Joshua Hamilton, director Lauren MacMullan, Nickelodeon Animation Studio, 2006. 03:14-03:53, 06:11-07:06.

the risk.

## The Kantian Account

Some people do not like this Utilitarian account. They think that it is missing something fundamental to the morality of actions. For example, Utilitarianism does not have a respect for personhood built in. Showing a person respect and dignity, for a Utilitarian, is seen as secondary. If you can get the best results while being respectful, do so, but the main driving force is to make the world a better place. This is where we get Kantianism, named after its inventor Immanuel Kant (1724–1804).<sup>131</sup> Kant was a Non-consequentialist, meaning that he held that the consequences have no bearing on the morality of actions. Kant’s ethical theory, also, is based around act-types rather than tokens. This puts Kantianism in direct opposition to Utilitarianism.

For Kant, and the Kantians, there are many moral imperatives, commands, which we need to follow to in order to act morally. These imperatives act categorically according to the type of action it is. For example, you have simple commands like “don’t make false statements,” “don’t cheat,” and “don’t steal.” For this theory, we don’t need to wait until the dust settles to know whether we did the right thing, rather we just need to know what kind of action it was and what the imperative associated with it is.

Determining the imperative associated with a given act-type can be quite difficult. Luckily, however, according to Kant, they all reduce to one simple command. This is the Categorical Imperative. Kant, himself, gives four different formulations of the Categorical Imperative and claims that they all mean essentially the same thing. There is no small debate about whether they do, but it is certainly clear that two of the four mean the

---

<sup>131</sup>Immanuel Kant, *Groundwork of the Metaphysics of Morals*, translated by Mark Gregor (Cambridge UP, 1998).

same as the other two, so the debate is really about whether the two formulations get at the same thing. That being said, the most relevant formulation to our question about the moral status of bloodbending is the Formula from Humanity.<sup>132</sup> This imperative tells us to treat all of humanity, ourselves included, as an end in and of themselves and never merely as a means to an end. Within that formulation, there is a central concept which we need to be clear on before we can use the theory to evaluate bloodbending. This is the idea of using someone as a mere-means.

We all have goals and aspirations which we seek to accomplish in our lives. For Aang, it is to master all four elements and bring balance to the world. To accomplish these goals, we often need to use each other by getting help or through normal exchanges, such as Aang using Katara, Toph, and Zuko to learn the different styles of bending. In these cases, we are using them as a means to our end but we are not necessarily using them merely as a means. We take the other person's goals and aspirations into account when we use them and, thereby, help them achieve their goals. Using someone merely as a means is a different matter entirely. In these cases, we do not take their goals into consideration, we trick, coerce, or force them to do something which they otherwise would not consent to. Kant believed that our autonomy (our free will) and our rationality were the most valuable things in the world. Any action which hindered or depreciated a person's autonomy or rationality, regardless of the consequences, is wrong. In lying, for example, we are intentionally deceiving a rational autonomous agent for the betterment of ourselves or another. This is using them merely as a means because we are not appreciating their great worth.

As with the Utilitarian account, Avatar has several examples which we could apply Kantian thinking to in order to make a moral judgement. In the very first episode of the series, Aang lies about being the Avatar. In doing so, Aang is misinforming

---

<sup>132</sup>Kant, *Groundwork of the Metaphysics of Morals* pg. 38.

Katara and Sokka for his benefit, as he later admits.<sup>133</sup> In such a simple case, we can see that Kant would say that Aang did something wrong. In a similar situation, in Book 1, Episode 4, Team Avatar arrives at the island of Kyoshi, named after one of Aang's past lives.<sup>134</sup> There, Aang chooses to be honest about being the Avatar. Despite the consequences of this, namely Zuko finding out and going to the island, Kantianism says that Aang did the right thing, because him lying would have been wrong.

For other examples, in Book 1, Episode 9, Team Avatar encounters a band of pirates who have, through "high risk trading," acquired a waterbending scroll, with forms and moves which Katara and Aang can learn to forward their quest to master the element.<sup>135</sup> As the story progresses, we learn that Katara stole the scroll from the pirates. Even though the scroll was stolen by them, stealing from a thief is still stealing. In doing so, Katara used them as a mere means and violated the Categorical Imperative. This means that Kant would say that her theft was morally wrong. The guilt for the action is on Katara, just as the initial theft of the scroll by the pirates is on the pirates. Similarly, in Book 3, Episode 5, a few Fire Nation soldiers spot Team Avatar and the proceed to send a message to the Fire Lord.<sup>136</sup> As the messenger hawk flies away, Combustion Man steals the scroll. Kant and those who think like him would say that Combustion Man did something wrong because

---

<sup>133</sup> "The Boy in the Iceberg," creator and writer Michael Dante DiMartino, director Dave Filoni, Nickelodeon Animation Studio, 2005. 11:31-11:52.

<sup>134</sup> "The Warriors of Kyoshi," creator Michael Dante DiMartino, director Giancarlo Volpe, writer Aaron Ehasz, Nickelodeon Animation Studio, 2005. 06:05-07:35.

<sup>135</sup> "The Waterbending Scroll," creator Michael Dante DiMartino, director Anthony Lioi, writer John O'Bryan, Nickelodeon Animation Studio, 2005. 07:04-10:00.

<sup>136</sup> "The Beach," creator Michael Dante DiMartino, author Katie Mattila, director Joaquim Dos Santos, Nickelodeon Animation Studio, 2007. 03:46-04:20, 08:09-08:56.

he is stealing and the Categorical Imperative strictly forbids theft.

## The Theories Applied to Bloodbending

As you may have guessed from my examples, both Kantianism and Utilitarianism truly shine when put to the test, used in cases, and the World of Avatar has no shortage of examples. For many of the moral cases in Avatar, Kantianism and Utilitarianism give remarkably different responses. In Book 2, Episode 3, in order to help the citizens of Omashu flee the Fire Nation invaders, Team Avatar and the populace use a small octopus-like creature to fake an epidemic, called ‘penta-pox’.<sup>137</sup> Doing so was an affective and non-violent way to get out of the city and Utilitarianism would approve. Kantianism, on the other hand, holds that any form of deception, which is a kind of lying, is wrong. In Book 3, Episode 1, Team Avatar and their cohorts are on a stolen Fire Nation ship.<sup>138</sup> The act of stealing this ship would be seen as fine according to Utilitarianism but be seen as morally wrong according to Kantianism. Similarly, in the same episode, they are impersonating Fire Nation soldiers.<sup>139</sup> This is deceitful and is lying. Kantianism gives a clear and absolute prohibition to lying, meaning that they say this is wrong. But Utilitarians disagree. These fundamental disagreements become even more pronounced when we apply the theories to the four cases of bloodbending seen in *The Puppetmaster*.

The, chronologically, first case of bloodbending is Hama using it to escape from prison. Hama is in prison after being captured by the Fire Nation during one of their raids. Hama,

---

<sup>137</sup> “Return to Omashu,” creator Michael Dante DiMartino, author Elizabeth Welch Ehasz, director Ethan Spaulding, Nickelodeon Animation Studio, 2006. 08:48-10:57.

<sup>138</sup> “The Awakening,” creator Michael Dante DiMartino, director Giancarlo Volpe, writer Aaron Ehasz, Nickelodeon Animation Studio, 2007. 06:58-08:00.

<sup>139</sup> *The Awakening* 08:46-10:00.



feeling the power of the full moon, realizes that living creatures, like elephant-rats and humans, have blood in their veins and that blood has water. Using her waterbending, Hama learns to manipulate that blood in the elephant rats and use them as puppets. This is not, however, enough to make her escape. On another full moon, Hama uses bloodbending to force a guard to open her cell and thereby escape from the prison.<sup>140</sup> The moral question which this scene raises and which these theories need to answer is whether it was permissible for Hama to use her skill to escape.

Looking at this prison escape, we see that Utilitarianism says that Hama did the right thing. In fact, she should have continued to use bloodbending to free the other Southern Water Tribe benders as well (which she may have). Doing so, in this case, results in the best outcome given her options, freedom to pursue her happiness and (in freeing the others) their happiness and, according to Utilitarianism, the right action is the one with the best consequences given your options. Many of us will likely agree with this assessment. Morally speaking, it was wrong for the Fire Nation to imprison those benders in the first place, so it seems only right to use her skills to escape.

Kantianism, on the other hand, holds that Hama did something wrong, regardless of the consequences. Hama bloodbends the guard of the prison and thereby forces him to release her. In doing this, she is using the guard merely as a means to an end (namely, escape from prison). She does not care about the goals or aspirations of the guard, rather she is forcing him to do something which he would not do otherwise. Kant and those who think like him would likely say that there is nothing wrong with her escaping from prison, rather it was wrong for her to use bloodbending to get do it. This is in stark contrast to Utilitarianism, which says that the results are all that matter, the steps used to get there are not relevant (unless there is a better way). Kant does not think that the ends justify the means, how

---

<sup>140</sup> *The Puppetmaster* 18:04-19:33.

you achieve a goal is just as important, if not more important, than what you achieved.

For the next case, we find Hama, significantly older, settled down in a Fire Nation Village. She is still a strong waterbender and still knows how to bloodbend. On full moon nights, Hama uses this ability to force people to leave their homes and walk deep into the forest, where she has a cave and she chains them to the walls, making them her prisoners.<sup>141</sup> Her motivation is vengeance, not just against the soldiers who imprisoned her or against the soldiers which raided her tribe, but rather against the entirety of the Fire Nation. In her mind, all Fire Nation citizens are guilty and worthy of her wrath. For this case, the questions which needs answering is whether it was permissible for Hama to use her skill to have vengeance in this way.

Utilitarianism would say that Hama kidnapping these citizens, regardless of the method used, is wrong. We likely will agree with this assessment too. Hama's actions are certainly causing more harm than good. The people of the village are terrified and those who she as imprisoned are experiencing similar pain and suffering to her own imprisonment years before. There are very little, if any, positive results from her kidnapping of the villagers. Since the suffering outweighs the happiness caused, Hama's actions are wrong, according to Utilitarianism. The Kantians agree with this assessment, but for different reasons. According to Kantianism, Hama is clearly compelling people to do something which they would not normally do, namely be chained up in a cave. She is not taking their goals and aspirations into account and is using them as a mere-means. We need to note that the pain, suffering, and emotional turmoil which Hama is causing does not enter into the Kantian framework. That would be consequentialist and Utilitarian thinking, which is not Kantian. Hama's actions here are wrong solely because of the violation of their autonomy.

In the third case, we have Hama and Katara using water-

---

<sup>141</sup> *The Puppetmaster* 15:12-16:49, 17:10-18:03.

bending to fight each other. Katara, a skilled waterbender in her own right, uses some of the skills which she learned from Hama to pull water from the surrounding plant life. Hama switches tactics and uses bloodbending to take control of Aang and Sokka's bodies, using them to fight for her against Katara. Hama then forces Sokka to raise his Space-sword and fly towards Aang.<sup>142</sup> If nothing is done, Aang will die. The question here is whether it was permissible for Hama to use bloodbending to force Sokka to attempt to kill Aang.

Utilitarianism gives a similar response to The Kidnapping, claiming that Hama was doing something wrong. This is because, had Hama succeeded, she would have killed the Avatar, which would have stopped him from defeating the Fire Lord and ending the war. Not only that, but it would have allowed the destruction of the Earth Kingdom at the Fire Lord's hands, because Aang would not have been there to stop it. All of the pain and suffering which Aang will have prevented would equally fall on Hama's shoulders. All of those extreme factors make for a very clear Utilitarian response; Hama was doing something wrong. This example, in particular, leads to a possible response. Hama was ignorant of these facts, does that enter into it? For Utilitarianism, a person's knowledge does not change the morality of an action. The Principle of Utility does not have room for the state of mind the doer is in. If a person is ignorant of the potential consequences of their actions, we may hold them less responsible, we may place less blame on them for their wrongdoing, but that does not make what they did any less wrong.

For this case, Kant would agree with the Utilitarians and say that Hama was in the wrong. Again, different reasoning is used. Killing a person (either yourself or another), especially for your own benefit, is the ultimate case of using them as a mere-means. In doing so, you are removing their rationality and autonomy from the world and, at the same time, violating

---

<sup>142</sup> *The Puppetmaster* 22:08-22:53.

that autonomy because you are removing all of their abilities. This reasoning, again, is very different than the kind which the Utilitarian would use. The Utilitarian does not think that killing a person is always wrong, rather they work on a case-by-case basis. In choosing whether to kill the Fire Lord, the Utilitarian would likely say that killing him would be a better choice than merely removing his bending, as there would be less net suffering in the world because of it. The Kantian would disagree, killing a person is always wrong.

For the final case, we pick up right where the previous left off. Sokka, sword raised, is flying towards Aang. Katara is the only one who can save them, and she only has one option, use bloodbending on Hama to stop her. This would save our heroes and subdue Hama. Katara, realizing this, uses bloodbending for the first time and defeats Hama, who is captured by her former prisoners and taken away. As she leaves, Hama says “my work is done. Congratulations, Katara. You’re a bloodbender.”<sup>143</sup> The ethical question which this case gives is whether it is permissible to use bloodbending to save a life.

The Kantians and the Utilitarians, yet again, disagree. Utilitarianism has it that Katara did the right thing. If she had not bloodbended Hama, The Avatar would have died. All of the good which Aang would cause in the world would not happen. In fact, in a sense, all of those positives act in favor of saving him, by any means. In bloodbending Hama, Katara is partially responsible for all of the good which Aang would cause thereafter. All of those factors make the response pretty clear. Katara may feel bad and cry because of what she had to do to save Aang’s life, but those feelings are a drop in the ocean compared to the alternatives. For this case, unlike The Attempt to Kill Aang, Katara knew all of those factors. Though this does not change the actual moral rectitude of the action, it should change how much praise or blame we place on her. Katara knowing this makes it so that we should not blame her

---

<sup>143</sup> *The Puppetmaster* 22:53-23:43.

for bloodbending, in fact, we should thank and congratulate her for it. If Katara were a good Utilitarian, she would feel good for saving Aang's life and not sadness for how she had to do it. Since Utilitarianism is a Consequentialist stance, it has that the ends justify the means. Bloodbending, in this case, is a means to saving the Avatar's life, which more than justifies its use.

The Kantians, on the other hand, will say that Katara using bloodbending to save Aang is wrong. This case mirrors very closely a thought experiment which Kant himself encountered. Suppose that your roommate is in the shower and you hear a knock at the door. You open it to find an axe-murderer, weapon out, asking about the whereabouts of your roommate. You have a choice. You could lie to save your friend or you could tell them the truth. In telling them the truth, they will die at the hands of the axe-murderer; but, Kant argued, you cannot lie to them. If you lie to the axe-murderer, you are using them merely as a means to an end and the results of that lie are on you, morally speaking. Whereas, if you tell the truth, you are not using them merely as a means and the results of their actions are not on you.<sup>144</sup> <sup>145</sup> This distinction could be generalized to a difference between killing and letting die. Killing a person is always wrong but letting a person to die at the hands of another (keeping your own hands clean, so to speak) is fine.<sup>146</sup> Katara's

---

<sup>144</sup>Immanuel Kant, "On a Supposed Right to Lie From Altruistic Motives," *Critique of Practical Reason, and Other Writings in Moral Philosophy*, edited and translated by Lewis White Beck (U of Chicago P, 1949) 346-350.

<sup>145</sup>There are more than 23 cases of lying or deception in the three seasons of the show and all would be counted as immoral according to Kantianism. There are 23 cases in the first two seasons (books) alone.

<sup>146</sup>A disagreement about killing vs letting die is even in one of the final episodes of Avatar. In Sozin's Comet Part 2, while Aang is contacting his past lives, he speaks with Avatar Kyoshi. Kyoshi describes her encounter with Chin the Conqueror, to which Aang replies "You didn't really kill Chin. Technically, he fell to his own doom because he was too stubborn to get out of the way." And then Kyoshi says "Personally, I don't really see

use of bloodbending is very similar to the axe-murderer case. In violating Hama's autonomy, the results of the action are on Katara and she did something morally wrong. If she had done nothing, on the other hand, the results of Aang's death would not be on her, morally speaking and she would not have done something wrong. This, again, points to a radically different way of thinking about morality. The Utilitarian does not see a distinction between killing and letting a person die, because both have the same results. The Utilitarian, also, would say that the results of Aang's death would, at least partially, be on Katara because she could have prevented it.

But, what do these theories say about violating a person's autonomy? In all four cases, we have situations where a bender is forcing another to do something which they would not otherwise do. In all four cases, we see a person's autonomy is being violated. Utilitarianism gives mixed results for these cases. As a result, this means that Utilitarianism gives us situations where a person's autonomy should be violated. Of course, most of the time, we should not force people or coerce them into doing things that they would not otherwise want, but, according to this theory, there are always exceptions. According to Utilitarianism, you should violate another's autonomy when your failure to do so would result in less than optimal results. Most of the time, violating a person's autonomy is not the best course of action, but there are going to be cases where it is what you need to do.

For all of these examples, Kantianism gives the same response, bloodbending is morally wrong. This means that the theory says that a person's autonomy should never be violated, through manipulation or through bloodbending. Could there be a case where Kantianism says that it is permissible? Such a case would require a person to voluntarily, in good faith and

---

the difference, but I assure you, I would have done whatever it took to stop Chin." Since Kyoshi does not see a difference between killing and letting die, we could, arguably, claim that she is more Utilitarian than Kantian. See *Sozin's Comet* 32:59-33:13

without deception, wish to be bloodbended. Even then, there could be a sense in which that person, in acting on that wish, is using themselves merely as a means to an end.

## Conclusion

In this paper, we have used the two major theories in Ethics to figure out the moral status of bloodbending. According to Utilitarianism, sometimes using the ability is wrong and other times it is the right thing to do. The evaluation has nothing to do with bloodbending itself, rather the judgement is made on the basis of the consequences of the action. The Utilitarian will claim that Hama escaping from prison was permissible, Hama kidnapping the villagers was wrong, Hama attempting to kill Aang was wrong, and Katara saving Aang's life was permissible. The Kantian, on the other hand, will make their evaluation based on the nature of bloodbending itself, regardless of the consequences. They claim that the prison escape was wrong, the kidnapping was wrong, attempting to kill Aang was wrong, and Katara saving the Avatar was wrong. In each case, the bender is violating a person's autonomy which is wrong. So, what is the moral status of bloodbending? The answer to this question depends on the answer to a bigger question: Which theory is more accurate, Utilitarianism or Kantianism?

## **MODULE IX**

### *What About Real World Cases?*



# *The Moral and Legal Status of Abortion by Mary Warren*

<sup>147</sup> We will be concerned with both the moral status of abortion, which for our purposes we may define as the act that a woman performs in voluntarily terminating, or allowing another person to terminate, her pregnancy, and the legal status that is appropriate for this act. I will argue that, while it is not possible to produce a satisfactory defense of a woman's right to obtain an abortion without showing that a fetus is not a human being, in the morally relevant sense of that term, we ought not to conclude that the difficulties involved in determining whether or not a fetus is human make it impossible to produce any satisfactory solution to the problem of the moral status of abortion. For it is possible to show that, on the basis of intuitions which we may expect even the opponents of abortion to share, *a fetus is not a person, hence not the sort of entity to which it is proper*

---

<sup>147</sup>Mary Anne Warren, "On the Moral and Legal Status of Abortion," *The Monist* vol. 57, no. 1, 1973, <https://doi.org/10.5840/monist197357133>, pp. 43–61.

*to ascribe full moral rights.*

Of course, while some philosophers would deny the possibility of any such proof,<sup>148</sup> others will deny that there is any need for it, since the moral permissibility of abortion appears to them to be too obvious to require proof. But the inadequacy of this attitude should be evident from the fact that both the friends and foes of abortion consider their position to be morally self-evident. Because proabortionists have never adequately come to grips with the conceptual issues surrounding abortion, most if not all, of the arguments which they advance in opposition to laws restricting access to abortion fail to refute or even weaken the traditional antiabortion argument, i.e., that a fetus is a human being, and therefore abortion is murder.

These arguments are typically of one of two sorts. Either they point to the terrible side effects of the restrictive laws, e.g., the deaths due to illegal abortions, and the fact that it is poor women who suffer the most as a result of these laws, or else they state that to deny a woman access to abortion is to deprive her of her right to control her own body. Unfortunately, however, the fact that restricting access to abortion has tragic side effects does not, in itself, show that the restrictions are unjustified, since murder is wrong regardless of the consequences of prohibiting it; and the appeal to the right to control one's body, which is generally construed as a property right, is at best a rather feeble argument for the permissibility of abortion. Mere ownership does not give me the right to kill innocent people whom I find on my property, and indeed I am apt to be held responsible if such people injure themselves while on my property. It is equally unclear that I have any moral right to expel an innocent person from my property when I know that doing so will result in his death.

---

<sup>148</sup>For example. Roger Wertheimer, who in "Understanding the Abortion Argument" (Philosophy and Public Affairs I:1) argues that the problem of the moral status of abortion is insoluble, in that the dispute over the status of the fetus is not a question of fact at all, but only a question of how one responds to the facts.

John Noonan is correct in saying that “the fundamental question in the long history of abortion is, How do you determine the humanity of a being?”.<sup>149</sup> He summarizes his own antiabortion argument, which is a version of the official position of the Catholic Church, as follows:<sup>150</sup>

... it is wrong to kill humans, however poor, weak, defenseless, and lacking in opportunity to develop their potential they may be. It is therefore morally wrong to kill infants. Similarly, it is morally wrong to kill embryos.

Noonan bases his claim that fetuses are human upon what he calls the theologians’ criterion of humanity: that whoever is conceived of human beings is human. But although he argues at length for the appropriateness of this criterion, he never questions the assumption that if a fetus is human then abortion is wrong for exactly the same reason that murder is wrong.

Judith Thomson is, in fact, the only writer I am aware of who has seriously questioned this assumption; she has argued that, even if we grant the antiabortionist his claim that a fetus is a human being, with the same right to life as any other human being, we can still demonstrate that, in at least some and perhaps most cases, a woman is under no moral obligation to complete an unwanted pregnancy.<sup>151</sup> Her argument is worth examining, since if it holds up it may enable us to establish the moral permissibility of abortion without becoming involved in problems about what entitles an entity to be considered human, and accorded full moral rights. To be able to do this would be a great gain in the power and simplicity of the proabortion position, since, although I will argue that these problems can be

---

<sup>149</sup> John Noonan, “Abortion and the Catholic Church: A Summary History,” *Natural Law Forum* vol. 12, 1967,

<sup>150</sup> John Noonan, “Deciding Who Is Human,” *Natural Law Forum* vol. 13, 1968,

<sup>151</sup> Judith Jarvis Thomson, “A Defense of Abortion,” *Philosophy & Public Affairs* vol. 1, no. 1, 1971, pp. 47–66.

salved at least as decisively as can any other moral problem, we should certainly be pleased to be able to avoid having to solve them as part of the justification of abortion.

On the other hand, even if Thomson's argument does not hold up, her insight, i.e., that it requires arguments to show that if fetuses are human then abortion is properly classified as murder, is an extremely valuable one. The assumption she attacks is particularly invidious, for it amounts to the decision that it is appropriate, in deciding the moral status of abortion, to leave the rights of the pregnant woman out of consideration entirely, except possibly when her life is threatened. Obviously, this will not do; determining what moral rights, if any, a fetus possesses is only the first step in determining the moral status of abortion. Step two, which is at least equally essential, is finding a just solution to the conflict between whatever rights the fetus may have, and the rights of the woman who is unwillingly pregnant. While the historical error has been to pay far too little attention to the second step, Thomson's suggestion is that if we look at the second step first, we may find that a woman has a right to obtain an abortion regardless of what rights the fetus has.

Our own inquiry will also have two stages. In Section I, we will consider whether or not it is possible to establish that abortion is morally permissible even on the assumption that a fetus is an entity with a full-fledged right to life. I will argue that in fact this cannot be established, at least not with the conclusiveness which is essential to our hopes of convincing those who are skeptical about the morality of abortion, and that we therefore cannot avoid dealing with the question of whether or not a fetus really does have the same right to life as a (more fully developed) human being.

In Section II, I will propose an answer to this question, namely, that a fetus cannot be considered a member of the moral community, the set of beings with full and equal moral rights, for the simple reason that it is not a person, and that it is personhood, and not genetic humanity, i.e., humanity as

defined by Noonan, which is the basis for membership in this community. I will argue that a fetus, whatever its stage of development, satisfies none of the basic criteria of personhood, and is not even enough like a person to be accorded even some of the same rights on the basis of this resemblance. Nor, as we will see, is a fetus's potential personhood a threat to the morality of abortion, since, whatever the rights of potential people may be, they are invariably overridden in any conflict with the moral rights of actual people.

## Part I

We turn now to Professor Thomson's case for the claim that even if a fetus has full moral rights, abortion is still morally permissible, at least sometimes, and for some reasons other than to save the woman's life. Her argument is based upon a clever, but I think faulty, thinking. She asked us to picture ourselves waking up one day, in bed with a famous violinist. Imagine that you have been kidnapped, and your bloodstream hooked up to that of the violinist, who happens to have an ailment that will certainly kill him unless he is permitted to share your kidneys for a period of nine months. No one else can save him, since you alone have the right type of blood. He will be unconscious all that time, and you will have to stay in bed with him, but after the nine months are over he may be unplugged, completely cured, that is provided that you have cooperated.

Now then, she continues, what are your obligations in this situation? The antiabortionist, if he is consistent, will have to say that you are obligated to stay in bed with the violinist: for all people have a right to life, and violinists are people, and therefore it would be murder for you to disconnect yourself from him and let him die.<sup>152</sup> But this is outrageous, and so there must be something wrong with the same argument when it is applied to abortion. It would certainly be commendable

---

<sup>152</sup>Thomson.

of you to agree to save the violinist, but it is absurd to suggest that your refusal to do so would be murder. His right to life does not obligate you to do whatever is required to keep him alive; nor does it justify anyone else forcing you to do so. A law that required you to stay in bed with the violinist would clearly be an unjust law, since it is no proper function of the law to force unwilling people to make huge sacrifice for the sake of other people toward whom they have no such prior obligation. Thomson concludes that, if this analogy is an apt one, then we can grant the antiabortionist his claim that a fetus is a human being, and still hold that it is at least sometimes the case that a pregnant woman has the right to refuse to be a Good Samaritan towards the fetus, i.e., to obtain an abortion. For there is a great gap between the claim that *x* has a right to life, and the claim that *y* is obligated to do whatever is necessary to keep *x* alive, let alone that he ought to be forced to do so. It is *y*'s duty to keep *x* alive only if he somehow contracted a special obligation to do so; a woman who is unwillingly pregnant, e.g., who was raped, has done nothing which obligates her to make the enormous sacrifice which is necessary to preserve the conceptus.

This argument is initially quite plausible, and in the extreme case of pregnancy due to rape, it is probably conclusive. Difficulties arise, however, when we try to specify more exactly the range of cases in which abortion is clearly justifiable even on the assumption that the fetus is human. Professor Thomson considers it a virtue of her argument that it does not enable us to conclude that abortion is always permissible. It would, she says, be "indecent" for a woman in seventh month to obtain an abortion just to avoid having to postpone a trip to Europe. On the other hand, her argument enables us to see that "a sick and desperately frightened schoolgirl pregnant due to rape may of course choose abortion, and that any law which rules this out is an insane law" (p. 65). So far, so good, but what are we to say about the woman who becomes pregnant not through rape but as a result of her own carelessness, or because of contraceptive failure, or who gets pregnant intentionally and then changes

her mind about wanting a child? With respect to such cases, the violinist analogy is of much less use to the defender of the woman's right to obtain an abortion.

Indeed, the choice of a pregnancy due to rape, as an example of a case in which abortion is permissible even if a fetus is considered a human being, is extremely significant; for it is only in the case of pregnancy due to rape that the woman's situation is adequately analogous to the violinist case for our intuitions about the latter to transfer convincingly. The crucial difference between a pregnancy due to rape and the normal case of an unwanted pregnancy is that in the normal case we cannot claim that the woman is in no way responsible for her predicament; she could have remained chaste, or taken her pills more faithfully or abstained on dangerous days, and so on. If on the other hand, you are kidnapped by strangers, and hooked up to a strange violinist, then you are free of any shred of responsibility for the situation, on the basis of which it would be argued that you are obligated to keep the violinist alive. Only when her pregnancy is due to rape is a woman clearly just as nonresponsible.<sup>153</sup>

Consequently, there is room for the antiabortionist to argue that in the normal case of unwanted pregnancy a woman has, by her own actions, assumed responsibility of the fetus. For if *x* behaves in a way which he could have avoided, and which he knows involves, let us say, a 1 percent chance of bringing into existence a human being, with a right to life, and does so knowing that if this should happen then that human being will perish unless *x* does certain things to keep him alive, then it is by no means clear that when it does happen *x* is free of any obligation to what he knew in advance would be required to

---

<sup>153</sup>We may safely ignore the fact that she might have avoided getting raped, e.g., by carrying a gun, since by similar means you might likewise have avoided getting kidnapped, and in neither case does the victim's failure to take all possible precautions against a highly unlikely event (as opposed to reasonable precautions against a rather likely event) mean that he is morally responsible for what happens.

keep that human being alive.

The plausibility of such an argument is enough to show that the Thomson analogy can provide a clear and persuasive defense of a woman's right to obtain an abortion only with respect to those cases in which the woman is in no way responsible for her pregnancy, e.g., where it is due to rape. In all other cases, we would almost certainly conclude that it was necessary to look carefully at the particular circumstances in order to determine the extent of the woman's responsibility and hence the extent of her obligation. This is an extremely unsatisfactory outcome, from the viewpoint of the opponents of restrictive abortion laws, most of whom are convinced that a woman has a right to obtain an abortion regardless of how and why she got pregnant.

Of course, a supporter of the violinist analogy might point out that it is absurd to suggest that forgetting her pill one day might be sufficient to obligate a woman to complete an unwanted pregnancy. And indeed, it is absurd to suggest this. As we will see, the moral right to obtain an abortion is not in the least dependent upon the extent to which a woman is responsible for her pregnancy. But unfortunately, once we allow the assumption that a fetus has full moral rights, we cannot avoid taking this absurd suggestion seriously. Perhaps we can make this point more clear by altering the violinist story just enough to make it more analogous to a normal unwanted pregnancy and less to a pregnancy due to rape, and then seeing whether it is still obvious that you are not obligated to stay in bed with the fellow.

Suppose, then, that violinists are peculiarly prone to the sort of illness the only cure for which is the use of someone else's bloodstream for nine months, and that because of this there has been formed a society of music lovers who agree that whenever a violinist is stricken they will draw lots and the loser will, by some means, be made the one and only person capable of saving him. Now then, would you be obligated to cooperate in curing the violinist if you had voluntarily joined this society, knowing the possible consequences, and then your name had been drawn



and you had been kidnapped? Admittedly, you did not promise ahead of time that you would, but you did deliberately place yourself in a position in which it might happen that a human life would be lost if you did not. Surely, this is at least a *prima facie* reason for supposing that you have an obligation to stay in bed with the violinist. Suppose that you had gotten your name drawn deliberately; surely that would be quite a strong reason for thinking that you had such an obligation.

It might be suggested that there is one important disanalogy between the modified violinist case and the case of an unwanted pregnancy, which makes the woman's responsibility significantly less, namely, the fact that the fetus comes into existence as the result of the woman's actions. This fact might give her a right to refuse to keep it alive, whereas she would not have had this right had it existed previously, independently, and then as a result of her actions become dependent upon her for its survival.

My own intuition, however, is that *x* has no more right to bring into existence, either deliberately or as a foreseeable result of actions he could have avoided, a being with full moral rights *y*, and then refuse to do what he knew beforehand would be required to keep that being alive, than he has to enter into an agreement with an existing person, whereby he may be called upon to save that person's life, and then refuse to do so when so called upon. Thus *x*'s responsibility for *y*'s existence does not seem to lessen his obligation to keep *y* alive, if he is also responsible for *y*'s being in a situation in which only he can save him.

Whether or not this intuition is entirely correct, it brings us back once again to the conclusion that once we allow the assumption that a fetus has full moral rights it becomes an extremely complex and difficult question whether and when abortion is justifiable. Thus the Thomson analogy cannot help us produce a clear and persuasive proof of the moral permissibility of abortion. Nor will the opponents of the restrictive laws thank us for anything less; for their conviction (for the most part) is

that abortion is obviously not a morally serious and extremely unfortunate, even though sometimes justified act, comparable to killing in self-defense or to letting the violinist die, but rather is closer to being a morally neutral act, like cutting one's hair.

The basis of this conviction, I believe, is the realization that a fetus is not a person, and thus does not have a full-fledged right to life. Perhaps the reason why this claim has been so inadequately defended is that it seems self-evident to those who accept it. And so it is, insofar as it follows from what I take to be perfectly obvious claims about the nature of personhood, and about the proper grounds for ascribing moral rights, claims which ought, indeed, to be obvious to both the friends and foes of abortion. Nevertheless, it is worth examining these claims, and showing how they demonstrate the moral innocuousness of abortion, since this apparently has not been adequately done before.

## Part II

The question which we must answer in order to produce a satisfactory solution to the problem of the moral status of abortion is this: How are we to define the moral community, the set of beings with full and equal moral rights, such that we can decide whether a human fetus is a member of this community or not? What sort of entity, exactly, has the inalienable rights to life, liberty, and the pursuit of happiness? Jefferson attributed these rights to all men ... If so, then we arrive, first, at Noonan's problem of defining what makes a being human, and, second, at the equally vital question which Noonan does not consider, namely, What reason is there for identifying the moral community with the set of all human beings, in whatever way we have chosen to define that term?

## 1. On the Definition of “Human”

One reason why this vital second question is so frequently overlooked in the debate over the moral status of abortion is that the term “human” has two distinct, but not often distinguished, senses. This fact results in a slide of meaning, which serves to conceal the fallaciousness of the traditional argument that since (1) it is wrong to kill innocent human beings, and (2) fetuses are innocent human beings, then (3) it is wrong to kill fetuses. For if “human” is used in the same sense in both (1) and (2) then, whichever of the two uses is meant, one of these premises is question-begging. And if it is used in two different senses then of course the conclusion doesn’t follow.

Thus, (1) is a self-evident moral truth,<sup>154</sup> and avoids begging the question about abortion, only if “human being” is used to mean something like “a full-fledged member of the moral community.” (It may or may not also be meant to refer exclusively to members of the species *Homo sapiens*.) We may call this the moral sense of “human.” It is not to be confused with what we will call the genetic sense; i.e., the sense in which a member of the species is a human being, and no member of any other species could be. If (1) is acceptable only if the moral sense is intended, (2) is non-question-begging only if what is intended is the genetic sense.

In “Deciding Who Is Human,” Noonan argues for the classification of fetuses with human beings by pointing to the presence of the full genetic code, and the potential capacity for rational thought (p. 135). It is clear that what he needs to show, for his version of the traditional argument to be valid, is that fetuses are human in the moral sense, the sense in which it is analytically true that all human beings have full moral rights. But, in the absence of any argument showing that whatever is

---

<sup>154</sup>Of course, the principle that it is (always) wrong to kill innocent human beings is in need of many other modifications, e.g., that it may be permissible to do so to save a greater number of other innocent human beings, but we may safely ignore these complications here.

genetically human is also morally human, and he gives none, nothing more than genetic humanity can be demonstrated by the presence of the human genetic code. And, as we will see, the potential capacity for rational thought can at most show that an entity has the potential for becoming human in the moral sense.

## 2. Defining the Moral Community

Can it be established that genetic humanity is sufficient for moral humanity? I think that there are very good reasons for not defining the moral community in way. I would like to suggest an alternative way of defining the moral community, which I will argue for only to the extent of explaining why it is, or should be, self-evident. The suggestion is simply that the moral community consists of all and only people, rather than all and only human beings;<sup>155</sup> and probably the best way of demonstrating its self-evidence is by considering the concept of personhood, to see what sorts of entity are and are not persons, and what the decision that a being is or is not a person implies about its moral rights.

What characteristics entitle an entity to be considered a person? This is obviously not the place to attempt a complete analysis of the concept of personhood, but we do not need such a fully adequate analysis just to determine whether and why a fetus is or isn't a person. All we need is a rough and approximate list of the most basic criteria of personhood, and some idea of which, or how many, of these an entity must satisfy in order to properly be considered a person. 28 In searching for such criteria, it is useful to look beyond the set of people with whom we are acquainted, and ask how we would decide whether a totally alien being was a person or not. (For we have no right

---

<sup>155</sup>From here on, we will use "human" to mean genetically human, since the moral sense seems closely connected to, and perhaps derived from, the assumption that genetic humanity is sufficient for membership in the moral community.

to assume that genetic humanity is necessary for personhood.) Imagine a space traveler who lands on an unknown planet and encounters a race of beings utterly unlike any he has ever seen or heard of. If he wants to be sure of behaving morally toward these beings, he has to somehow decide whether they are people, and hence have full moral rights, or whether they are the sort of thing which he need not feel guilty about treating as, for example, a source of food.

How should he go about making this decision? If he has some anthropological background, he might look for such things as religion, art, and the manufacturing of tools, weapons, or shelters, since these factors have been used to distinguish our human from our prehuman ancestors, in what seems to be closer to the moral than the genetic sense of "human." And no doubt he would be right to consider the presence of such factors as good evidence that the alien beings were people, and morally human. It would, however, be overly anthropocentric of him to take the absence of these things as adequate evidence that they were not, since we can imagine people who have progressed beyond, or evolved without ever developing these cultural characteristics.

I suggest that the traits which are most central to the concept of personhood, or humanity' in the moral sense, are, very roughly; the following:

- 1 consciousness (of objects and events external and/or internal to the being), and in particular the capacity to feel pain;
- 2 reasoning (the developed capacity to solve new and relatively complex problems);
- 3 self-motivated activity (activity which is relatively independent of either genetic or direct external control);
- 4 the capacity to communicate, by whatever means, messages of an indefinite variety of types, that is, not just

with an indefinite number of possible contents, but on indefinitely many possible topics;

- 5 the presence of self-concepts, and self-awareness, either individual or racial, or both.

Admittedly, there are apt to be a great many problems involved in formulating precise definitions of these criteria, let alone in developing universally valid behavioral criteria for deciding when they apply. But I will assume that both we and our explorer know approximately what (1)-(5) mean, and that he is also able to determine whether or not they apply. How, then, should he use his findings to decide whether or not the alien beings are people? We needn't suppose that an entity must have all of these attributes to be properly considered a person; (1) and (2) alone may well be sufficient for personhood, and quite probably (1)-(3), if "activity" is construed so as to include the activity of reasoning.

All we need to claim, to demonstrate that a fetus is not a person, is that any being which satisfies none of (1)-(5) is certainly not a person. I consider this claim to be so obvious that I think anyone who denied it, and claimed that a being which satisfied none of (1)-(5) was a person all the same, would thereby demonstrate that he had no notion at all of what a person is—perhaps because he had confused the concept of a person with that of genetic humanity. If the opponents of abortion were to deny the appropriateness of these five criteria, I do not know what further arguments would convince them. We would probably have to admit that our conceptual schemes were indeed irreconcilably different, and that our dispute could not be settled objectively.

I do not expect this to happen, however, since I think that the concept of a person is one which is very nearly universal (to people), and that it is common to both proabortionists and antiabortionists, even though neither group has fully realized the relevance of this concept to the resolution of their dispute.

Furthermore, I think that on reflection even the antiabortionists ought to agree not only that (1) - (5) are central to the concept of personhood, but also that it is a part of this concept that all and only people have full moral rights. The concept of a person is in part a moral concept; once we have admitted that *x* is a person we have recognized, even if we have not agreed to respect, *x*'s right to be treated as a member of the moral community. It is true that the claim that *x* is a human being is more commonly voiced as part of an appeal to treat *x* decently than is the claim that *x* is a person, but this is either because "human being" is here used in the sense which implies personhood, or because the genetic and moral sense of "human" have been confused.

Now if (1)-(5) are indeed the primary criteria of personhood, then it is clear that genetic humanity is neither necessary nor sufficient for establishing that an entity is a person. Some human beings are not people, and there may well be people who are not human beings. A man or woman whose consciousness has been permanently obliterated but who remains alive is a human being which is no longer a person; defective human beings, with no appreciable mental capacity, are not and presumably never will be people; and a fetus is a human being which is not yet a person, and which therefore cannot coherently be said to have full moral rights. Citizens of the next century should be prepared to recognize highly advanced, self-aware robots or computers, should such be developed, and intelligent inhabitants of other worlds, should such be found, as people in the fullest sense, and to respect their moral rights. But to ascribe full moral rights to an entity which is not a person is as absurd as to ascribe moral obligations and responsibilities to such an entity.

### 3. Fetal Development and the Right to Life

Two problems arise in the application of these suggestions for the definition of the moral community to the determination

of the precise moral status of a human fetus. Given that the paradigm example of a person is a normal adult being, then (1) How like this paradigm, in particular how far advanced since conception, does a human being need to be before it begins to have a right to life by virtue, not of being fully a person as of vet, but of being like a person? and (2) To what extent, if any does the fact that a fetus has the potential for becoming a person endow it with some of the same rights? Each of these questions requires some comment.

In answering the first question, we need not attempt a detailed consideration of the moral rights of organisms which are not developed enough, aware enough, intelligent enough, etc., to be considered people, but which resemble people in some respects. It does seem reasonable to suggest that the more like a person, in the relevant respects, a being is, the stronger is the case for regarding it as having a right to life, and indeed the stronger its right to life is. Thus we ought to take seriously the suggestion that, insofar as “the human individual develops biologically in a continuous fashion ... the rights of a human person might develop in the same way”.<sup>156</sup> But we must keep in mind that the attributes which are relevant in determining whether or not an entity is enough like a person to be regarded as having some of the same moral rights are no different from those which are relevant to determining whether or not it is fully a person—i.e., are no different from (1)-(5)—and that being genetically human, or having recognizably human facial and other physical features, or detectable brain activity, or the capacity to survive outside the uterus, are simply not among these relevant attributes.

Thus it is clear that even though a seven- or eight-month fetus has features which make it apt to arouse in us almost the same powerful protective instinct as is commonly aroused by a small infant, nevertheless it is not significantly more personlike than is a very small embryo. It is somewhat more personlike; it

---

<sup>156</sup>Thomas L. Hayes, “A Biological View,” *Commonweal* vol. 85,



can apparently feel and respond to pain, and it may even have a rudimentary form of consciousness, insofar as its brain is quite active. Nevertheless, it seems safe to say that it is not fully conscious, in the way that an infant of a few months is, and that it cannot reason, or communicate messages of indefinitely many sorts, does not engage in self-motivated activity; and has no self-awareness. Thus, in the relevant respects, a fetus, even a fully developed one, is considerably less personlike than is the average mature mammal, indeed the average fish. And I think that a rational person must conclude that if the right to life of a fetus is to be based upon its resemblance to a person, then it cannot be said to have any more right to life than, let us say, a newborn guppy (which also seems to be capable of feeling pain), and that a right of that magnitude could never override a woman's right to obtain an abortion, at any stage of her pregnancy.

There may, of course, be other arguments in favor of placing legal limits upon the stage of pregnancy in which an abortion may be performed. Given the relative safety of the new techniques of artificially inducing labor during the third trimester, the danger to the woman's life or health is no longer such an argument. Neither is the fact that people tend to respond to the thought of abortion in the later stages of pregnancy with emotional repulsion, since mere emotional responses cannot take the place of moral reasoning in determining what ought to be permitted. Nor, finally, is the frequently heard argument that legalizing abortion, especially late in the pregnancy, may erode the level of respect for human life, leading, perhaps, to an increase in unjustified euthanasia and other crimes. For this threat, if it is a threat, can be better met by educating people to the kinds of moral distinctions which we are making here than by limiting access to abortion (which limitation may, in its disregard for the rights of women, be just as damaging to the level of respect for human rights).

Thus, since the fact that even a fully developed fetus is not personlike enough to have any significant right to life on

the basis of its personlikeness shows that no legal restrictions upon the stage of pregnancy in which an abortion may be performed can be justified on the grounds that we should protect the rights of the older fetus. And once there is no other apparent justification for such restrictions, we may conclude that they are entirely unjustified. Whether or not it would be indecent (whatever that means) for a woman in her seventh month to obtain an abortion just to avoid having a to postpone a trip to Europe, it would not, in itself, be immoral, and therefore it ought to be permitted.

#### 4. Potential Personhood and the Right to Life

We have seen that a fetus does not resemble a person in any way that can support the claim that it has even some of the same rights. But what about its potential, the fact that if nurtured and allowed to develop naturally it will very probably become a person? Doesn't that alone give it at least some right to life? It is hard to deny that the fact that an entity is a potential person is a strong *prima facie* reason for not destroying it, but we need not conclude from this that a potential person has a right to life, by virtue of that potential. It may be that our feeling that it is better, other things being equal, not to destroy a potential person is better explained by the fact that potential people are still (felt to be) an invaluable resource, not to be lightly squandered. Surely, if every speck of dust were a potential person, we would be much less apt to conclude that every potential person has a right to become actual.

Still, we do not need to insist that a potential person has no right to life whatever. There may well be something immoral, and not just imprudent, about wantonly destroying potential people, when doing so isn't necessary to protect anyone's rights. But even if a potential person does have some *prima facie* right to life, such a right could not possibly outweigh the right of a woman to obtain an abortion, since the rights of any actual person invariably outweigh those of any potential person, whenever

the two conflict. Since this may not be immediately obvious in the case of a human fetus, let us look at another case.

Suppose that our space explorer falls into the hands of an alien culture, whose scientists decide to create a few hundred thousand or more human beings, by breaking his body into its component cells, and using these to create fully developed human beings, with, of course, his genetic code. We may imagine that each of these newly created men will have all of the original man's abilities, skills, knowledge, and so on, and also have an individual self-concept, in short that each of them will be a *bona fide* (though hardly unique) person. Imagine that the whole project will take only seconds, and that its chances of success are extremely high, and that our explorer knows all of this, and also knows that these people will be treated fairly. I maintain that in such a situation he would have every right to escape if he could, and thus to deprive all of these potential people of their potential lives; for his right to life outweighs all of theirs together, in spite of the fact that they are all genetically human, all innocent, and all have a very high probability of becoming people very soon, if only he refrains from acting.

Indeed, I think he would have a right to escape even if it were not his life which the alien scientists planned to take, but only a year of his freedom, or, indeed, only a day. Nor would he be obligated to stay if he had gotten captured (thus bringing all these people-potentials into existence) because of his own carelessness, or even if he had done so deliberately knowing the consequences. Regardless of how he got captured, he is not morally obligated to remain in captivity for any period of time for the sake of permitting any number of potential people to come into actuality, so great is the margin by which one actual person's right to liberty outweighs whatever right to life even a hundred thousand potential people have. And it seems reasonable to conclude that the rights of a woman will outweigh by a similar margin whatever right to life a fetus may have by virtue of its potential personhood.

Thus, neither a fetus's resemblance to a person, nor its po-

tential for becoming a person, provides any basis whatsoever for the claim that it has any significant right to life. Consequently, a woman's right to protect her health, happiness, freedom, and even her life,<sup>157</sup> by terminating an unwanted pregnancy will always override whatever right to life it may be appropriate to ascribe to a fetus, even a fully developed one. And thus, in the absence of any overwhelming social need for every possible child, the laws which restrict the right to obtain an abortion, or limit the period of pregnancy during which an abortion maybe performed, are a wholly unjustified violation of a woman's most basic moral and constitutional rights.<sup>158</sup>

## Postscript on Infanticide, February 26, 1982

One of the most troubling objections to the argument presented in this article is that it may appear to justify not only abortion but infanticide as well. A newborn infant is not a great deal more personlike than a ninemonth fetus, and thus it might seem that if late-term abortion is sometimes justified, then infanticide must also be sometimes justified. Yet most people consider that infanticide is a form of murder, and thus never justified.

While it is important to appreciate the emotional force of this objection, its logical force is far less than it may seem at first glance. There are many reasons why infanticide is much more difficult to justify than abortion, even though if my argument is correct neither constitutes the killing of a person. In this country, and in this period of history, the deliberate killing of viable newborns is virtually never justified. This is in part because neonates are so very close to being persons that

---

<sup>157</sup>That is, insofar as the death rate, for the woman, is higher for childbirth than for early abortion.

<sup>158</sup>My thanks to the following people, who were kind enough to read and criticize an earlier version of this paper: Herbert Gold, Gene Glass, Anne Lauterbach, Judith Thomson, Mary Mothersill, and Timothy Binkley.

to kill them requires a very strong moral justification as does the killing of dolphins, whales, chimpanzees, and other highly personlike creatures. It is certainly wrong to kill such beings just for the sake of convenience, or financial profit, or "sport."

Another reason why infanticide is usually wrong, in our society, is that if the newborn's parents do not want it, or are unable to care for it, there are (in most cases) people who are able and eager to adopt it and to provide a good home for it. Many people wait years for the opportunity to adopt a child, and some are unable to do so even though there is every reason to believe that they would be good parents. The needless destruction of a viable infant inevitably deprives some person or persons of a source of great pleasure and satisfaction, perhaps severely impoverishing their lives. Furthermore, even if an infant is considered to be adoptable (e.g., because of some extremely severe mental or physical handicap) it is still wrong in most cases to kill it. For most of us value the lives of infants, and would prefer to pay taxes to support orphanages and state institutions for the handicapped rather than to allow unwanted infants to be killed. So long as most people feel this way, and so long as our society can afford to provide care for infants which are unwanted or which have special needs that preclude home care, it is wrong to destroy any infant which has a chance of living a reasonably satisfactory life.

If these arguments show that infanticide is wrong, at least in this society, then why don't they also show that late-term abortion is wrong? After all, third trimester fetuses are also highly personlike, and many people value them and would much prefer that they be preserved; even at some cost to themselves. As a potential source of pleasure to some family, a viable fetus is just as valuable as a viable infant. But there is an obvious and crucial difference between the two cases: once the infant is born, its continued life cannot (except, perhaps, in very exceptional cases) pose any serious threat to the woman's life or health, since she is free to put it up for adoption, or, where this is impossible, to place it in a state-supported institution. While

she might prefer that it die, rather than being raised by others, it is not clear that such a preference would constitute a right on her part. True, she may suffer greatly from the knowledge that her child will be thrown into the lottery of the adoption system, and that she will be unable to ensure its well-being, or even to know whether it is healthy, happy, doing well in school, etc.: for the law generally does not permit natural parents to remain in contact with their children, once they are adopted by another family. But there are surely better ways of dealing with these problems than by permitting infanticide in such cases. (It might help, for instance, if the natural parents of adopted children could at least receive some information about their progress, without necessarily being informed of the identity of the adopting family.)

In contrast, a pregnant woman's right to protect her own life and health clearly outweighs other people's desire that the fetus be preserved—just as, when a person's life or limb is threatened by some wild animal, and when the threat cannot be removed without killing the animal, the person's right to self-protection outweighs the desires of those who would prefer that the animal not be harmed. Thus, while the moment of birth may not mark any sharp discontinuity in the degree to which an infant possesses a right to life, it does mark the end of the mother's absolute right to determine its fate. Indeed, if and when a late-term abortion could be safely performed without killing the fetus, she would have no absolute right to insist on its death (e.g., if others wish to adopt it or pay for its care), for the same reason that she does not have a right to insist that a viable infant be killed.

It remains true that according to my argument neither abortion nor the killing of neonates is properly considered a form of murder. Perhaps it is understandable that the law should classify infanticide as murder or homicide, since there is no other existing legal category which adequately or conveniently expresses the force of our society's disapproval of this action. But the moral distinction remains, and it has several important

consequences.

In the first place, it implies that when an infant is born into a society which-unlike ours-is so impoverished that it simply cannot care for it adequately without endangering the survival of existing persons, killing it or allowing it to die is not necessarily wrong-provided that there is no other society which is willing and able to provide such care. Most human societies, from those at the hunting and gathering stage of economic development to the highly civilized Greeks and Romans, have permitted the practice of infanticide under such unfortunate circumstances, and I would argue that it shows a serious lack of understanding to condemn them as morally backward for this reason alone.

In the second place, the argument implies that when an infant is born with such severe physical anomalies that its life would predictably be a very short and/or very miserable one, even with the most heroic of medical treatment, and where its parents do not choose to bear the often crushing emotional, financial and other burdens attendant upon the artificial prolongation of such a tragic life, it is not morally wrong to cease or withhold treatment, thus allowing the infant a painless death. It is wrong (and sometimes a form of murder) to practice involuntary euthanasia on persons, since they have the right to decide for themselves whether or not they wish to continue to live. But terminally ill neonates cannot make this decision for themselves, and thus it is incumbent upon responsible persons to make the decision for them, as best they can. The mistaken belief that infanticide is always tantamount to murder is responsible for a great deal of unnecessary suffering, not just on the part of infants which are made to endure needlessly prolonged and painful deaths, but also on the part of parents, nurses, and other involved persons, who must watch infants suffering needlessly, helpless to end that suffering in the most humane way.

I am well aware that these conclusions, however modest and reasonable they may seem to some people, strike other people as morally monstrous, and that some people might even prefer to abandon their previous support for women's right to abor-

tion rather than accept a theory which leads to such conclusions about infanticide. But all that these facts show is that abortion is not an isolated moral issue; to fully understand the moral status of abortion we may have to reconsider other moral issues as well, issues not just about infanticide and euthanasia, but also about the moral rights of women and of nonhuman animals. It is a philosopher's task to criticize mistaken beliefs which stand in the way of moral understanding, even when-perhaps especially when-those beliefs are popular and widespread. The belief that moral strictures against killing should apply equally to all genetically human entities, and only to genetically human entities, is such an error. The overcoming of this error will undoubtedly require long and often painful struggle; but it must be done.



## *Part 20: What is an Abortion?*

This question is the start of a very heated debate which can get nasty, so please keep that in mind. I have used other heated examples before, so you should know how to handle this. For this chunk, we are just looking at what it is, not the moral status of it. Here, we are looking at the metaphysical question concerning abortion, what is it? Later, once this is settled, we will look at the ethical question, is it OK to have one?

One essential feature to an abortion, it would seem, is that there needs to be the ending of a pregnancy. But, that certainly can't be it. Take this case as an example:

Birth is the ending of a pregnancy.

By our definition, the ending of a pregnancy is an abortion.

Therefore, birth is an abortion.

So, we could add in something about the pre-mature nature of the termination, but that would make pre-mature births abortions, which also seems just as wrong (as in misfitting). Glossing over some of the more graphic examples I could give, the core, missing feature which makes an act an abortion and not birth or some crime against another person seems to be that

it needs to be voluntary. The woman needs to, with informed consent, want to terminate the pregnancy early, without resulting in a child. There can be interesting cases, worth thinking about, where the woman gives consent, but not informed consent (she may have been misinformed about what exactly it entails, which might make her not want it). For this module, we will be defining an abortion as one of these two things (could be both, but that's a weird case), these two features fit for both the pro-choice side and the pro-life side of the debate, we will be covering both:

1. A woman voluntarily terminating her own pregnancy.
2. A woman allowing another to terminate her pregnancy (referring to the subject).

If you only have the first one, then you will not get cases of, say, doctor assisting the woman in terminating her pregnancy. If you only have the second one, then you will not get cases of self-administered abortions. To avoid the cases where birth could be defined as an abortion, we need to say that 'terminating a pregnancy' does involve the ending of a fetus. The moral status of that fetus is where the debate is.

### **Is it morally permissible to have an abortion?**

I get that this is a hot issue, and if I have not already, I guess that I will need to be far more active in the comments in the discussion for this one than I already have been, please remember to be civil. For Warren, the moral status of abortion hinges on the answer to the following question:

The Fetus Question: Is a fetus a person, in the morally relevant sense?

The main tie-in, and one which you will read me say several times, is that if a fetus is a person, then abortion is wrong,

if a fetus is not a person, then abortion is permissible. The Fetus Question moves the ethical debate regarding abortion, "is abortion murder?" to a metaphysical debate regarding person-hood. Typically, our moral intuitions are gut responses, which come from mental shortcuts, when we analyse that shortcut, we can get down to a metaphysical question which we have grounds to prove or disprove. The author of the reading is arguing that it is possible to show that a fetus is not a person, making abortion permissible.<sup>159</sup>

## Framing the Problem

Of course, while some philosophers and others would deny the possibility of a proof that a fetus is not a person, claiming that to do so would be to prove a contradiction or that it's not possible to prove either way (this would lead to skepticism about fetal person-hood). By the same token, others will claim that there's no need for a proof. These people claim that the moral status of a fetus, its person-hood, is too obvious to need justification. But, both sides of the debate, pro-choice and the pro-life, take their evidence and reasoning to be obvious, to an equal degree. This is much like a belief in God. Some Atheists claim that the non-existence of God is clear and obvious, while at the same time, Theists claim that the existence of God is equally obvious. This disagreement means that we can't trust our gut instincts on this issue, we need to use logic to show that one of the sides is faulty.

Through this module, we will see the best arguments on both sides of the abortion debate. But, to start us off, we will look at the commonplace, pro-choice arguments. Though

---

<sup>159</sup>Some students, especially those who use translation dictionaries, will have issues with two different words, these words are "person" and "human". These do not mean the same thing, and we will see how these come apart later. For now, a human is a member of our species and a person is a being with moral worth equal to you or me. We will see later how there can be non-human persons.

Warren agrees with their conclusions, she disagrees with how they got there. There are glaring issues in their reasoning.

#### Argument A

Restrictive Laws regarding abortion cause more harm than the lack of those laws.

Causing more harm than otherwise is always wrong.

Therefore, restrictive laws regarding abortions are wrong.

#### Argument B

Restrictive laws regarding abortion deny women the ability to control their reproduction.

Denying women the right to control their reproduction violates their right to control their body.

Violating their right to control their body is always wrong.

Therefore, restrictive laws regarding abortions are wrong.

Pro-choicers (I made that term up) have never adequately come to grips with the conceptual issues surrounding abortion. As a result, their arguments miss the mark when they try to attack the pro-life side. Their arguments avoid the fight rather than engaging in it. You can think of it as the pro-lifers are in a castle, and the pro-choicers are attacking it, but their catapults always miss. Most, if not all, of the arguments which they give in favor of legal abortions fail to refute or even weaken the traditional pro-life argument. This is that a fetus is a human being and, therefore, abortion is murder.

The pro-choice arguments tend to fall into two different categories. First, the arguments use consequentialist/utilitarian style reasoning to show that having abortions be illegal is wrong. This is exemplified by Argument A. For

example, these arguments point to the terrible results of having the restrictive laws. These include things like the deaths caused by unsafe abortions, the fact that they unfairly result in harder hardship on poorer women, the fact that the lack of access results in emotional hardship, and so on. But, the pro-life side has an easy reply to this. They can claim that the tragic side effects don't, by themselves, show that the laws aren't justified (they can still be justified even with the results). This is where we can get these arguments:

The Pro-Life Response to Argument A:

- 1 A fetus is a being worthy of our moral consideration, same as you or me.
- 2 If a fetus is a being worthy of our moral consideration, then abortion is murder.
- 3 Murder is always wrong (regardless of the consequences).
- 4 So, from (1) and (2), abortion is murder.
- 5 Therefore, from (3) and (4), abortion is always wrong (regardless of the consequences).

The Pro-Life Response to Argument B:

- 1 Violating a person's rights is always wrong.
- 2 A person's rights extend so far as they do not violate another's stronger right.
- 3 A person's right to life is stronger than another person's bodily rights.
- 4 Having an abortion (a claimed case of bodily right) is violating a fetus' right to life (as a fetus is a person).
- 5 Therefore, having an abortion is always wrong.

The Pro-Life Response to Argument A, in their eyes, takes out the reasoning for the second line of Argument A. This is basically showing that there are some cases where the moral status of some behavior is not determined by the consequences. This can be supported by the very definition of murder, which

is wrongful intentional killing. This kind of argument falls into the non-consequentialist style thinking and the basis there is that something are wrong regardless of the consequences. This will show up several times, the pro-life side of the debate tends to give reasons rooted in non-consequentialist style thinking. There is another pro-choice reply to this, which debates the idea that the abortion is murder in this case, further deepening the divide between the intuitions (because the reasonable response is more strongly consequentialist)

The second argument given by the pro-choicers is exemplified by Argument B. This one uses a more non-consequentialist/Kantian style reasoning, pointing out that denying a woman the ability to have an abortion is to deprive her of some manner of bodily right. For example, the right to choose when and how one bears young. This one also falls short. The Pro-Life Response to Argument B shows where this falls short. They are basically saying there that while a person has rights, those rights can't imply that it's OK to violate another's rights. For example, take property rights. It seems clear that a person has the right to remove another from their property and has the right to defend their property, but how far does that right extend? Take this example, which really did happen, but I have exaggerated for our purposes:

A man owns a large piece of property, land, and there's a group of young hooligans who ride their motor bikes on the trails through his land. So, one day, he put up a line of piano wire across one of the trails, at a particular height. As one of the young people rides through the trail, the wire, tight, catches them on the neck and decapitates them.

This was wrong of him, or so many have argued, and the pro-lifers can use the sort of reasoning here to support themselves. In putting up the piano wire and killing the hooligan, the man

violated that person's right to life. This right is stronger than the property rights which the man would have otherwise had, meaning that in this case, he did not have the right to protect his property in this way. From this line of thought, the pro-lifer can say that, because a person's right to life is stronger than a person's bodily rights, abortion is still wrong. Since we have these strong competing intuitions and because the pro-lifer, the serious ones at least, won't give an inch for the consequentialist considerations, we are going to need to approach the abortion debate from the non-consequentialist perspective and show that the system, in fact, allows for abortion.

### **The Fundamental Question**

The most basic question which we need to answer is not about the results of having legal abortions but rather about what it takes to be a person. This is the Fetus Question. The pro-life responses to the commonplace pro-choice arguments show that they take a fetus to be a person. So, there are two routes we can take. First, we could show that there are some cases such that, even if we assume that a fetus is a person, abortion is still permissible (this is the first part of the pro-choice response). Second, we could show that a fetus is not a person, thereby making abortion permissible (before a certain point). So, for that second half, we need to show what features it takes to be a person. If a fetus has the features, then it is a person and abortion is murder, if it does not, then it is not. Warren's case is that there are certain cases where it's permissible to have an abortion, even if a fetus is a person, and then she moves on to show that a fetus is not a person, which entails that abortion is permissible.

# *Part 21: The Abortion Debate (Pro-Choice)*

## **Assume That a Fetus is a Person**

As I mentioned at the end of the previous page, Warren has two parts to her paper. First, she will assume, just for the sake of argument, that a fetus is a person, and then from that show that if a fetus is a person, there are cases where killing it is permissible. The second section is where she shows that a fetus is not a person, so it's not entitled to the same moral rights as you or me, which means that abortion is permissible. This page is the start of that first section. If there's a way to get that at least some abortions are OK in this case, then you can't have the all out ban on them which is some times proposed. Rather, morally speaking, you would need to have some exceptions. It is worth noting, and we will return to this point, that the morality of an abortion, proved in this section, assuming that a fetus is a person, is limited to a very select range of cases. This range of cases is defined, roughly, by how much consent the woman had in the actions resulting in pregnancy. This is where Warren, through Judith Thomson, gets the Violinist Thought



Experiment:<sup>160</sup>

Imagine that you have been kidnapped, and your blood-stream hooked up to that of the violinist, who happens to have an ailment that will certainly kill him unless he is permitted to share your kidneys for a period of nine months. You are a human dialysis machine. No one else can save him, since you alone have the right type of blood. He will be unconscious all that time, and you will have to stay in bed with him, but after the nine months are over he may be unplugged, completely cured, that is provided that you have cooperated. The violinist themselves had no knowledge that this would happen to them.

A common point stated about this thought experiment is that it says that you are stuck in bed for the 9-months. While this is not true in most cases of pregnancy (as in the woman can move around), there are plenty of cases of pregnancy where this is the case (as in they are stuck in bed for most of it), especially if the woman is quite small (my mom, for example, is 4' 10") and the father is quite large (my father, for example, is 6' 2"). But, moving on, if the person on the pro-life side of this debate is consistent in their beliefs, if they don't have any contradictions in their reasoning, then they will need to say that you would need to go to term and be there for the full 9-months. Despite you being forced into the situation, you will need to keep the violinist alive. They come to this from the following reasoning, which is much like their response to Argument B in the previous part:

---

<sup>160</sup>Thomson.

## The Pro-Life Response to Argument B:

- 0 (hidden line) A fetus is a person.
- 1 Violating a person's rights is always wrong.
- 2 A person's rights extend so far as they do not violate another's stronger right.
- 3 A person's right to life is stronger than another person's bodily rights.
- 4 Having an abortion (a claimed case of bodily right) is violating a fetus' right to life (as a fetus is a person).
- 5 Therefore, having an abortion is always wrong.

## The Pro-Life Case to Stay Plugged In:

- 0 (hidden line) A violinist is a person.
- 1 Violating a person's rights is always wrong.
- 2 A person's rights extend so far as they do not violate another's stronger right.
- 3 A person's right to life is stronger than another person's bodily rights.
- 4 Unplugging from the violinist is violating the violinist's right to life.
- 5 Therefore, unplugging from the violinist is always wrong.

This, as we have seen before, is very strongly non-consequentialist style thinking. But, the vast majority of people would think that it's outrageous to think that there's the moral obligation here to keep the violinist alive. The claim here, roughly, boils down to the idea that the rights of another can't force a person to go above and beyond, take extreme measures, to ensure it. The violinist case shows that there are some cases where a person's bodily rights are stronger than a person's right to life. If there are cases like this for pregnancy, then the pro-lifer, morally, can't hold their position absolutely. It's really good of a person to agree to take on such a sacrifice, especially it was thrust upon them like this, but it seems wrong

to say that your refusal is murder. Though he certainly has the right to life, something about this case must be off, removing the obligation. His right to life, in this case, does not force you, morally speaking, to keep him alive by what ever means necessary; nor does it justify anyone else forcing you to do so. A law that required you to stay in bed with the violinist would clearly be an unjust law, since it is no proper function of the law to force unwilling people to make huge sacrifice for the sake of other people toward whom they have no such prior obligation. The key feature, for the Violinist case, is that you did not give informed consent to be plugged into the violinist.

### **What does this case get us?**

Well, it does get us something to get started on. There are a few similarities and differences between this case and pregnancies. First, we have a person (assuming that a fetus is a person) who is dependent on another for survival. In both cases, if the aware party does certain actions, then the other will die. The other aspect is that the dependency causes a drain on the aware party. In this case, there's a sense in which the other does not have a moral obligation to keep them alive. But, what removed that obligation? Some would say, as Warren does, that the key feature which negates the obligation is the kidnapping aspect. The person did not knowingly and voluntarily enter into this arrangement. They did not consent to taking on the risk. If a woman does not enter into this arrangement willingly or without knowing the risks (without informed consent), then her situation is sufficiently similar to the violinist's case.

### **The Results of the Violinist Case and the Problems**

The Violinist Thought-Experiment is initially quite plausible. It gives us a grounding to have that we aren't always obligated to keep people alive by any means necessary. If there are cases

where a pregnancy is sufficiently like this case, then we can get the permissibility of abortion in those cases. But, for cases where they aren't relevantly similar to the Violinist case, we don't get the windy-side of morality, necessarily. The only real cases which are sufficiently like the violinist case are cases of pregnancy due to rape. In those cases, the woman did not voluntarily take on the risks. But, there could be some vagueness on how much give the Violinist Case gives us. If we extend it too far from the bounds of the Violinist case, we could run into wrinkles. For example, take this case:

A woman, who is 7-months pregnant, finds herself unable to travel to Europe because of the pregnancy. She really does not want to postpone the trip. But, if she has an abortion, then the trip will go off without a hitch. Is it permissible for her to have an abortion?

There seems to be a relevant difference between this case and the violinist case. For this one, people will often make a few different claims. First, some will claim that the woman is too far into the game to quit now, saying that the time for the abortion has passed. Others might claim that the trip to Europe is not a good enough reason to want an abortion, the case just doesn't make her bodily right strong enough. And others still will say that (assuming this is not a case of rape) that she entered into this knowing the risk and has the obligation.

In the case of pregnancies not caused by rape, there are other things which the woman could have done to prevent her. In other cases, there are some things which the woman could have done. These are Warren's examples, so if they aren't correct or in some way off, be mad at her. First, she could have remained chaste. In other words, she could have denied her partner the relations. If the partner acted anyway, this would be rape and fall into the violinist case. Remember, informed consent is absolutely key. Her second option, if she chooses to

have sex, is to have taken her pills more faithfully. I know from the life experiences of friends, family, and former students, that this is hardly a 100% sure-fire way to prevent pregnancy as it's often believed to be. Personally, I am in favor of the development of the male-birth control, as it's better to take the bullets out of the gun than put on a bullet proof vest. The third option, if all else fails, is for her to abstain on dangerous days. But this option, also, is not reliable as some might claim. I encourage all people to research these options, but only get your research from non-religious, non-abstinence only, scientific sources.

### **The Pro-Life Response**

Consequently, there is room for the antiabortionist to argue that in the normal case of unwanted pregnancy a woman has, by her own actions, assumed responsibility of the fetus.

If x behaves in a way which he could have avoided, and which he knows involves a 1% chance of bringing into existence a human being, with a right to life, and does so knowing that if this should happen then that human being will perish unless x does certain things to keep him alive, then, when it does happen, x is not free of any obligation to what he knew in advance would he required to keep that human being alive.

To make this into a case, something which we can imagine and use for the analogy, I have created this thought-experiment, based on the Violinist Case, Violinist Cult Thought Experiment:

Suppose that you are a member of a cult along with 99 other people. I know that cult has a negative stigma to it, but bear with me. All of you have voluntarily and with full reasonable consent, entered into a lottery where one of you will be chosen at random to take on the role of being this violinist's human dialysis machine. Imagine that your name is drawn and you have the violinist hooked up. What's your obligation like now?

Most people, from my experience, say that in this case, there is the obligation to keep the violinist alive. You signed up knowing the risks and you lost the lottery, so to speak. So, what's the difference between the violinist case and the violinist cult case? The first seems to remove the obligation, but the second seems to have it.

### **Restricting the Outcome**

The plausibility of such an argument is enough to show that the Violinist analogy can provide a clear and persuasive defense of a woman's right to obtain an abortion only when the woman is in no way responsible for her pregnancy. In all other cases, we would almost certainly conclude that it was necessary to look carefully at the particular circumstances in order to determine the extent of the woman's responsibility and hence the extent of her obligation.

## **Prove That a Fetus is Not a Person**

Is a fetus a person?

As I have mentioned before, the second section of this paper concerns whether or not a fetus is a person. The point of the previous section was to show that there are some cases where abortion is OK, even if a fetus is a person. The point here is to

show that a fetus is not a person, which means that abortion isn't murder, and therefore is not wrong. This is where the Fetus Question comes in very strong, this is why I also noted that we need to distinguish between 'person' and 'human'. Questions regarding personhood are metaphysical questions, does a thing have certain features? Similarly, questions regarding 'human-hood' are metaphysical questions. From this, as I have mentioned, we are able to move from an ethical question to a metaphysical one. So, let's look at the standard, non-religious, argument from the pro-life side and another argument, which doesn't look similar, but I will explain how these relate:

The Standard Pro-Life Argument

It is wrong to kill innocent human beings  
Fetuses are innocent human beings  
Therefore, it is wrong to kill fetuses

The Cheese Sandwich Fallacy

Nothing is better than God.  
A cheese sandwich is better than nothing.  
Therefore, a cheese sandwich is better than God.

These two arguments might look completely different, but both of them fall into the same logical fallacy, equivocation. Remember, I mentioned that there's a distinction between 'human' and 'person'. With that in mind, it becomes clear that there's something wrong with the Standard Pro-Life Argument. Glossing over that distinction leads to the equivocation. The Cheese Sandwich Fallacy is a great example of this fallacy. An equivocation is where a person uses the same word in two different ways in an argument, this is meant to mislead the reader. The Standard Pro-Life Argument has the same error as this one. Now, we aren't the Dark Brotherhood, so the equivocation is not in the word 'innocent'. Rather, the phrase "human being" is being used in two different ways.

In the sentence "fetuses are innocent human beings", the term 'human being' is being used to talk about a member of our species, and up until this point, I have been very consistent in the use of the term 'human' to talk about members of our species, things with the same sort of genetic make up as us. If we use this species interpretation of human, which has a long and solid history, we see that this is correct, fetuses are humans.

The other time we encounter this term is in the first line "it's wrong to kill innocent human beings." If we use the case we use the genetic, or species, interpretation of the phrase 'human being', we can quickly find cases where this is wrong, even by the non-consequentialist's lights. For example, if a brain dead human has a living will saying that they should pull the plug after a few days. But, if this was said without any context, we would likely accept it, so what makes it different? Well, the use of 'human being' in this sentence, normally, means, in the most generous interpretation, where the line makes sense, "a full-fledged member of the moral community." We may call this the moral sense of "human" and I have been using the term 'person' to demark this.<sup>161</sup> In general, even when we are dealing with certain interesting legal cases, this distinction between 'human' and 'person' is overlooked. For example, a person has rights, but a human may or may not have rights. Though I don't like this example personally, but corporate person-hood is an example of this. We have a non-human entity, a corporation, seen as a person. Now, I would argue that 'corporations are people' is a legal fiction, it's not actual. But, it's certainly possible, as we will see later, that there are actual, non-fictitious, non-human persons. Having a clear distinction between these in our everyday speech makes many thorny moral questions disappear, or,

---

<sup>161</sup>There's a certain idiosyncratic grammar which I have the habit of using in the case of the word 'person'. In English, there are, it would seem, two acceptable plurals, 'people' and 'persons'. How I use them, 'people' refers to a collection of persons and 'persons' refers to beings with person-hood individually. 'People' is a collective or group sense of the plural and 'persons' is a more individualistic sense.



at the very least, makes them more intelligible. If we remove the equivocations in both of the arguments we get the following:

The Standard Pro-Life Argument

It is wrong to kill innocent persons  
Fetuses are innocent human beings  
Therefore, it is wrong to kill fetuses

The Cheese Sandwich Fallacy

No existing thing is better than God.  
A cheese sandwich is better than not having anything at all.  
Therefore, a cheese sandwich is better than God.

As you can see, this doesn't work. But, if we have that all humans are persons, as in that being a generic human is enough to be a person, the argument would work. This would be to say that all humans are persons. On the other hand, however, if there are cases where a human is not a person, then the argument fails. If we can show that all fetuses (before a certain point of development) and not persons, then we have that abortion is permissible (before a certain point in development), by the pro-life style reasoning. Showing that no fetus (before that point) is a person is Warren's next step.

## Warren's Criteria for Personhood

Warren argues that there are 5 features which are needed to be a person, and if a fetus has these features (after a certain point), then abortion would be murder (after that point). These features are listed here:

1. Consciousness
2. Reasoning

3. Self-Motivated Activity
4. Communication
5. Self-Awareness

## **Feature 1: Consciousness**

The first of these features seem to be the most intuitive. This is consciousness. We have touched on this before, in the Mind-Body Problem module. Although there is much debate about some of the features of consciousness, we do have a general understanding of when it's had and when it's not. For example, does it react to external stimulus? Is there some behavioral or other kind of evidence that shows that this thing is thinking, has an internal life?

### **Is this thing conscious?**

There are a few tests to tell whether a being is conscious. The relevant one here awareness (as in reaction to external stimulus) and evidence of internal thoughts. Many other tests include the self-awareness, which is not necessarily the same thing, but that one is another feature. We will limited this test to merely something like "does it feel pain?". If it reacts, it's conscious.

### **Are we the only conscious things?**

There are several creatures which have consciousness (by this definition) aside from humans. It is worth noting also that not all humans are conscious. Consciousness in humans is only there when the human is developed beyond a certain point and without certain impairments. In fact, most animals do have this and so do most fish. It's possible for some plants to even have this, though that is easily debated against.

### **Why does this matter?**

If Warren is correct, several beings, including some humans, are immediately excluded from person-hood. Plants are (more than likely) taken out of the moral community, severely disabled humans, clams, and some animals. Even with this feature alone, it could be argued that fetuses aren't persons (before a certain point, we will see this later). One way to think about this is that you take the set of all things out there, and then slowly add criteria to whittle the total down to just persons. It is important to realize that even in international law, 'human' is not the same as 'person'. Framing the question in this way moves it out of ethics and into metaphysics. So, "is abortion permissible?" is an ethical question, while "is a fetus a person?" is a metaphysical one. The answer to the second gives us the answer to the first.

### **Feature 2: Reasoning**

The second feature which seems to be necessary for person-hood according to Warren is reasoning. This is the developed capacity to solve new and relatively complex problems. This, too, is not found in all humans. Most people might think that this is a necessary part of consciousness, so it should not have its own section. But this is not quite true. Consciousness, as we are using it, is having a 'what-it's-like'-ness. Having sensations. A being can feel pleasure and pain without having the ability to reason. It also should be noted that we are worried about the mental capacity to solve the problems, not the physical ability. Infant humans, if they are not disabled to certain degrees, have the mental ability to solve the problems, but not the physical ability (strength, dexterity, so on).

### **Does this thing have reasoning?**

There are some basic tests to tell whether a being has reasoning. These tests are likely going to be more involved than merely

trying to tell whether the thing can feel pain/pleasure. The task is to give the creature a puzzle and see whether it can solve the puzzle. For example, in the case of a raven, put some food just out of its reach and see whether it can come up with a way to get the food.

### **Are we the only reasoning creatures?**

Just like with consciousness, we also find this capacity in many non-human animals (chimpanzees are an easy example, same with dolphins). We also find this capacity in some fish (octopuses are a great example). In general, if the being can figure something out, or at least shows that it's thinking about a problem in a more abstract way, then we can say that it has reasoning. But, it's also true that some humans lack this feature. There are some which are severely mentally handicapped, those in the later stages of dementia, and so on. These humans are certainly human but they are not, according to Warren, persons.

### **Why does this matter?**

As before, if reasoning is an essential part to being a person, we can further whittle down our list of potential persons. As before, humans developed to a certain point have reasoning, but before that we don't. So, fetuses don't count here if they are prior to a certain stage of development. Also, many non-human animals do stay in the list. For example, we have more complexly intelligent animals and some fish (octopuses, cetaceans, new world monkeys, apes, chimpanzees, bears, otters and so on, basically, if you can train them, they have this), but plants are now certainly out.

### **Feature 3: Self-Motivated Activity**

The third feature is a bit more restrictive. Self-motivated activity is closely tied with reasoning and consciousness. This is

activity which is relatively independent of either genetic or direct external control. Some may think that this requires some kind of libertarian free will, which seems to require some kind of soul. Warren, however, gets around this worry by adding in that it's independent of 'direct' external control. This allows for indirect external control to have a part in it. To see the difference, direct external control would be a case where a mad scientist puts a microchip in your brain and controls you with a remote. This would not make you free, your actions would not come from you at all. On the other hand, indirect external control would be the sort of thing which you, more than likely, are experiencing right now. You have been heavily influenced by the past and your experiences, your choices are dictated by those, they are not necessarily instinctual. Both the hard determinists and the compatibilists would be fine with saying that actions can be independent of direct external control, but both will claim that they are indirectly controlled, by the past and the laws of nature. The hard determinist, however, would not be cool with the idea of morality, however, they would say that (although it's there) it doesn't count.

### **Is this thing 'free'?**

In general, the tests which we apply to figure out whether a being has reasoning will apply here. We could call those tests a 'two for one'. When it comes to the tests for reasoning, we are asking whether they can solve puzzles, and to even engage with a puzzle to solve it, without really strange external factors being included (such as, being a remote-controlled robot), requires you to have self-motivated activity.

There are potential ways for isolating this feature and testing only it. For example, we would need to make a scenario where the creature (human, animal, robot) is denied external motivation for acting, it would have no instinctual reason to act. If the creature still engages in the activity, then it would be self motivated. For example, a spy-camera in your house watching

your pet. If we see that the pet acts without direct external interaction, then we could say that the action is self-motivated.

### **Are we the only 'free' things?**

Some, like Descartes and Kant, will claim that humans are the only creatures which can be free, and even some humans (non-person humans) lack this. This mostly stems from their Libertarian Free Will intuitions. But, if we limit the scope of self-motivation to something within the range of a compatibilist, then we have that there's no reason to think we are the only ones with it. In the case of Descartes, as we will see later, he claimed that animals lacked a soul, so (as a consequence, though not the one he was shooting for), non-human animals can't have self-motivated activity.

As with the previous, we see again that we aren't the only beings which count as having self-motivated activity. Much of the same beings from the reasoning section remain, but this is mostly because I don't know of a way to tell that a being has reasoning without getting that it has self-motivated activity.

We also have that some humans lack self-motivated activity. This list is much the same as before. Fetuses, yet again, lack this feature prior to a certain point in development.

### **Why does this matter?**

Intuitively, it seems clear that for a being to have moral rights and be in the moral community (be a person), they would need to be able to act. Reasoning and consciousness can only get you so far. We also need assurance that these beings are acting freely. Without this, it just does not seem to have the kind of weight needed. If we say that a being is a person, then they must have moral responsibility, to at least some degree, which gets us, by definition, self-motivated activity.

Similarly to the previous two, there are some humans which lack this ability. For example, the severely disabled or humans

prior to a certain point in development. But there are certain other, non-human, animals which have this feature. In fact, it could be argued that most animals have this and fish. As before, also, some great examples are chimpanzees, dolphins, various new world monkeys, and octopuses. As with the previous two, fetuses, prior to a certain point, lack these.

### **Feature 4: Communication**

This fourth feature is where person-hood becomes far more restrictive for Warren. Communication is the ability to express messages, by whatever means (not just vocal, but signing counts, and so do other methods), messages of various types and with a large variation of contents.

Warren goes, I think, a little too far in her definition of communication; claiming "by whatever means, messages of an indefinite variety of types, that is, not just with an indefinite number of possible contents, but on indefinitely many possible topics." This seems very extreme to me, and I doubt that I even qualify here. But, limiting it to as I have said above helps. This criterion, as she phrases it, would mean that I would need to be able to talk about anything and be able to do so in indefinite number of ways, which I think is just not possible, my brain is just not that big.

### **Can this thing communicate?**

The real discrepancy here, and the reason Warren makes such and extreme standard for her communication, is that we don't want it to be just simple messages. The messages need to be more complicated than, for example, the chemical trails which ants leave or the dancing movements of bees. Rather these messages need to convey complicated information and they need to be able to convey it in several different ways. The real point is to raise the bar on how smart the being needs to be to make the cut for personhood. The tests here are going to be a little more

relative to the kind of creature which we are testing. For example, with some particularly primitive human languages, the messages might not be able to be expressed in several different ways, but the speakers can learn the different ways (though the older members will have more difficulty). The basic test would be to watch the beings interact with each other and notice the kinds of messages they are able to understand and convey to each other. Is there a grammar? Is the communication structure learned or instinctual? Can they understand abstract concepts?

### Are we the only talkers?

Despite what my friends who study linguistics might think (claiming that humans are the only language-users), when it comes to communication, we are far from the only ones. Animal communication is very wide-spread, with, I would argue, the complexity necessary for personhood. For example, dolphins have communication, and this is not instinctual but learned, we have even figured out some words in at least one of the variety of languages spoken.<sup>162</sup> Chimpanzees have this capacity, but it does not seem to have one naturally arising. Some new world monkeys certainly have this capacity and have a naturally arising and learned languages (my personal favorite example is the cotton-top tamarin).<sup>163</sup> It may be also the case that octopuses are in this category. If we treat this as a standard for intelligence, then individuals with the mental capacity to communicate, but not the physical ability, would qualify. That being said, some humans lack the mental capacity to communicate, not just the physical ability. This can be due to a variety of

---

<sup>162</sup>TEDtalksDirector. Could we speak the language of dolphins? — Denise Herzing. *YouTube*, June 2013. [www.youtube.com/watch?v=CQ5dRyyHwfM](http://www.youtube.com/watch?v=CQ5dRyyHwfM).

<sup>163</sup>TEDEducation. How to speak monkey: The language of cotton-top tamarins - anne savage. *YouTube*, June 2014. [www.youtube.com/watch?v=4Vfn5CV9juI](http://www.youtube.com/watch?v=4Vfn5CV9juI).



reasons. And, as it applies to the relevant topic, fetuses before a certain point, lack this capacity all together.

### **Why does this matter?**

In general, the ability to communicate our thoughts and intentions is the biggest sign of intelligence but it's also a bench-mark for degrees of intelligence. My father, who is mono-lingual, has often claimed that speaking multiple languages is a sign that the person is really smart, encouraging me to learn several (which I have). Though I have met very smart mono-lingual people and not-so-smart multilingual people, the ability to express oneself is a fairly intuitive standard. Similarly, if a being which had this mental capacity and later lost it, we often, in the real world, hold them to a different moral standard. For example, a person with sever mental disabilities is not held as responsible for their actions as a person without those disabilities, even if they behave the same way (in this case). And, it seems that they should not be held to those standards. This criterion limits the scope of person-hood yet again. Sure fish can reason, but they certainly aren't smart enough to be persons. Similarly, cats and dogs are in the same boat. So, the examples I have been giving thus far remain as potential non-human persons, but some are excluded. Yet again, fetuses are not included in this list.

### **Feature 5: Self-Awareness**

This is the fifth and final feature of person-hood for Warren. This is self-awareness. This one, one could think, should be earlier than communication, as there are many creatures which have this but lack communication. To be self-aware one must have self-concepts, understand that they are a different being from others. You are an individual, not a 'hive-mind'. You can think about you, what you want.

### **Is this thing self-aware?**

This is where I give the tests which one can use to tell whether a being is self-aware. In the sciences, the commonly used test is to show whether the being can recognize themselves in a mirror, know that the reflection is not another being, then they have this self-awareness. But this method is flawed in several ways. There are some creatures who clearly have self-concepts but lack interest in mirrors. In the previous feature, I gave cotton-top tamarins as being talkers, and there's some solid evidence that the nature of their communication does require self-concepts, but they have failed the mirror test (when tested at least once, but there may have been something wrong with the testing perimeters). For example, when I was paying for my community college, o so long ago, I was working in Dementia Care. In trying to put an elderly woman to bed, she saw herself in a mirror. She called out to the reflection, asking it to leave her room and consistently looking back and trying to get it to leave. Eventually, I had to cover up the mirror. But, it's clear that, though her dementia was extreme enough for her not to have the same moral privileges as you or I, she still had self-concepts. So, though the mirror test works in most cases, we need to be careful about calling it a definitive test. A more holistic testing model is appropriate, so that we can weed out false positives and negatives.

### **Are we the only ones self-aware?**

It is certainly not the case that we are the only ones self-aware. Other creatures certainly are too, even the ones which can't communicate. For example, various cetaceans (whales and dolphins), primates, and some other creatures, such as some new world monkeys. Some octopuses have failed to show that they are self-aware in a way which I am willing to accept, namely in the presence of a mirror, they behaved as if there was another octopus present. But the jury is out on this for more intelligent

species of octopus. In general, if a being is able to communicate to the degree necessary for person-hood, then it's going to have this feature, but not the other way around. It's worth noting that not all humans are self-aware. We have the extreme cases of humans in irreversible comas or humans born without certain portions of the brain. Fetuses, yet again, lack this feature.

### **Why does it matter?**

As with the other standards, this raises the bar on what it takes to be a full-fledged person. Lacking the presence of self-concepts does not entail the sort of moral duties to them which we would place on ourselves for beings with these concepts. Persons are individuals, we have duties to them as individuals. If something can't identify as an individual, then we don't have the same duties to it.

### **An Interesting Tangent**

As we have seen through our analysis of these features which make a being a person, there are some humans which are not persons. Warren herself does not go down this rabbit hole, but using her requirements for person-hood gives us a very interesting tangent:

Is it possible for a non-human to be a person?

This question is very interesting for the discussions of Animal Rights. To have rights, to have aspects which others have a moral duty to ensure, you certainly need to be a person. If you are a person, then you have rights. Though it is possible for something to have rights and not be a person in the fullest sense, but those rights would be limited. Applying Warren's 5 criteria to other creatures in the world, we see that there are non-fictional, real, actual, non-human persons on the planet right now. Now, I am not an Area 51 conspiracy theories, saying that their are aliens in the base. But, looking at these, we

have some fascinating arguments to show that humans aren't the only persons. Proving and establishing legally that these beings are non-human persons will give animal rights activists a strong argument and a more powerful footing in making their case. As persons, these creatures will need to be given the same moral consideration as you or me. Below, I will give the 5 (five) categories of creatures where at least some of them have personhood. For some, the category will have more than one example in them. The first are the least contentious and the last are the most contentious.

### **Primates and (some) New-World Monkeys**

Of course, humans are primates, but we are human, so not exactly non-human persons. But we do find the capacities for the aspects of personhood in the non-human members. To start, they clearly react to external stimulation, so there's something going on upstairs, that's the first box ticked. Second, they can solve reasonably complex puzzles and even can make tools, so that's a second aspect met. Third, through their ability to solve puzzles and makes choices, we can see that their actions are not merely instinctual, they do have a moral compass, so to speak. Fourth, though they don't have a naturally arising language, they can be taught it and will use it even when not prompted to speak with each other if they know that the other will understand. Also, once the primate understands a language, it will teach its children the skill, with them even, sometimes, inventing new words. For more information on this, check out Washoe the Chimp. And fifth, primates, by and large, do have the ability to recognize themselves in the mirror and they do have self-directed thoughts and awareness. All of these features make it so that at least some primates, not just humans, are persons and, from that reasoning, deserve the same moral consideration as we would a human with the same mental capacity.

When it comes to new-world monkeys, we have a very interesting case, these are the cotton top tamarins. These little guys

are quite amazing. Not only do they have all of the features of person-hood, like the primates, but they have an added aspect. They have their own, learned, naturally arising language. Their language was not taught to them by us. They react to external stuff, and so forth. The only area which requires proving is whether they have self-concepts, but their language aspects seem to show that they would.

## Cetaceans

These are your dolphins and some whales. At present, some countries have recognized these creatures as non-human persons and have granted them various rights, even though they did not follow the same kind of reasoning given here. Dolphins, in particular, do have all of these features. First off, they react to external stimuli, which is going to be true for all animals worth mentioning. Second, they can solve problems and learn from each other. For example, one pod of dolphins independently learned to use a sponge to root the sea floor and others have learned this behavior from them. Third, their behaviors are not always instinctual and are independent of direct external control. For example, they play and will even engage in behaviors contrary to what we would think they would instinctively. Fourth, they do have communication, and a rather sophisticated one. The language is learned by the children, meaning that different pods might not understand another and we have even learned aspects of it able to communicate with them. Similarly, they are able to learn more than one language. Dolphins which we have trained will learn the whistle patterns for various tricks (like a dog), but will also mimic them to attempt at communication with other dolphins and even understand that the whistles in different orders will mean different things. The fifth aspect is self-awareness. This can be seen in both how they will pass the mirror test and also how they have naming customs for their young and how they introduce themselves. These make some dolphins non-human persons and worthy of

our protection.

### **Extraterrestrials and Advanced AI**

Both of these are at the bottom merely because their existence is controversial. No, I do not think that such beings are currently on the planet, rather they are worthy of mention because of their possibility. The section regarding the Mind-Body Problem shows us that proving person-hood for an AI will be tough, it will need to have Strong AI, if they are possible at all. However, a Strong AI would obviously be able to speak, have consciousness, self-awareness, self-motivated activity, and reasoning to the same degree as human persons, if they are possible. Not recognizing that an AI machine has reached the level of person-hood is the base-line for pretty much every robot-uprising Sci-Fi.

Extraterrestrials will likely have an easier time proving their person-hood than the machines. This is because they will have likely came from a process much like the one humans did. But, it would not surprise me at all if, on the day the first contact is made, there's a group out there who think that they aren't persons, not worthy of our consideration, because they have built into it the idea that person-hood is exclusively human. This, it would seem, would be mistaken.

# *I Was Once a Fetus by Alexander Pruss*

164

## INTRODUCTION

I am going to give an argument showing that abortion is wrong in exactly the same circumstances in which it is wrong to kill an adult. To argue further that abortion is always wrong would require showing that it is always wrong to kill an adult or that the circumstances in which it is not wrong—say, capital punishment—never befall a fetus. Such an argument will be beyond the scope of this paper, but since it is uncontroversial that it is wrong to kill an adult human being for the sorts of reasons for which most abortions are performed, it follows that most abortions are wrong. The argument has three parts, of decreasing difficulty. The most difficult will be the first part where I will argue that I was once a fetus and before that I was an embryo. This argument will rest on simple considerations of the metaphysics of identity. The next part of the argument will

---

<sup>164</sup>Alexander R. Pruss, “I Was Once a Fetus: An Identity-Based Argument Against Abortion.”

be to show that it would have been at least as wrong to have killed me before I was born as it would be to kill me now. I will argue for this in more than one way, but the guiding intuition is clear: if you kill me earlier, the victim is the same but the harm is greater since I am deprived of more the earlier I die. Finally, the easiest part of the argument will be that I am not relevantly different from anybody else and the fetus that I was was not relevantly different from any other human fetus, and so the argument applies equally well to all fetuses. The advantage of this argument over others is that it avoids talking of personhood, except in one of the several independent arguments in part two.

## 1. I WAS ONCE A FETUS

The first part seems innocuous. After all, is it not biologically evident that first I was an embryo, then a fetus, then a neonate, then an infant, then a toddler, then a child, then an adolescent, and then an adult? Does not my mother talk of the time when she was “pregnant with me” and thereby imply that it was I who was in her womb when she was pregnant? Is not the sonogram of my daughter the sonogram of that daughter of mine who will be born? Evident as it might be that I was once a fetus and given how clear it will be that abortion is wrong if I was once a fetus, it is obvious, however, that the opponent will have to focus his attack on this part of the argument. So more needs to be said.

About thirty years ago, nine months before I was born, a conception occurred. A sperm from my father fertilized an ovum from my mother. Within twenty-four hours, or sooner, a new organism came into existence, an organism that was neither a part of my mother nor of my father. For one, this organism was genetically distinct from both. For another, this organism’s functioning was directed towards its own benefit—selfishly, the organism colonized the womb, released hormones that trigger



changes in the woman beneficial to the organism, and so on. It certainly did not behave like a body part of either my mother or my father. Moreover, it clearly was not a part of my father—it need no longer have had any interaction with him. But it could not really be a part of my mother since the genetic contribution from my father was equal to that from the mother, so it was either a part of both or of neither. Thus, indeed, it was not a part of either. Besides, we can see that in the earliest days of this organism before implantation, the organism floated free, independently seeking nutrition in my mother's womb. This organism certainly was not a part of my mother.

Hence, we have on the scene a new individual organism, one that did not exist before. Let's give this organism a name: call it Bob. If we have a camera and look at what was happening in the womb in which Bob is living, we will see an embryo developing, cells differentiating, a fetus forming, growing, and finally a birth. If we keep watching, we see a neonate, then an infant, then a toddler, then a child, then an adolescent and then an adult. It's all a continuous history. But recall what I am out to prove. I am out to prove that I was once a fetus and indeed an embryo against an opponent who will not grant this. My opponent will thus have to deny that I and Bob are one and the same entity. He will have to say that "Bob" and "Alex" name two different entities, rather than being two names for one and the same entity at different stages of its life.

In any case, we have on the scene Bob the embryo. And then all this development happens. I now need a simple metaphysical principle. If an organism that once existed has never died, then this organism still exists. I will not argue for this principle. Someone who thinks that something can exist at time A and not exist at a later time B, without having ceased to exist in between, is beyond the reach of argument. The crucial question now is: Has Bob the embryo ever died? This is a question to which the biologists can tell us the answer. Bob's cells have divided, differentiated, and Bob has developed. But nowhere in the continuous history just described have we seen anything

we could identify as “the death of Bob.” In fact, the whole process is the very opposite of the process of death: we have a process of growth. That embryo that was conceived nine months before my birth never died. True, it ceased to be an embryo, and at the end of the nine months it ceased to be a fetus. But this is no more a literal death than my passing from childhood to adolescence or from adolescence to adulthood was.

Indeed, if Bob died, we would be mystified as to when he died. All we have in its life history is a process of growth and development. Now, it is true that not all deaths are alike—not all deaths involve an evident destruction. For instance, some philosophers think that the right way to describe an amoeba’s splitting is to say that the original amoeba dies and from its ashes there arise two new amoebae. Likewise, some philosophers think that when two entities merge into a single unified entity, the original entities perish and a new one is formed. That in fact may be how we should understand the process of conception: the egg and sperm perish, and a new thing results. But again, for as long as Bob has existed, he has always, in fact, been a single unified organism, and nothing like that happened in Bob’s life history—Bob never split in two and never merged with anything else so as to lose its own identity. If I were an identical twin, matters would be slightly different as an argument could then be made that the pre-twinning embryo has indeed perished when it split in two. But that’s not what happened to Bob. It is clear that Bob has not died in the prosaic way of having his organic functioning disrupted, and hasn’t even died in these two more outré ways that philosophers discuss.

Furthermore, the very continuity in Bob’s development speaks against the hypothesis that he died. When did that momentous event happen? When did Bob cease to exist? Could there have been some moment in Bob’s growth where one millisecond Bob was alive and a millisecond later Bob was no longer around? Surely not.

Therefore, it is sufficiently established that Bob, that em-

bryo who came into existence nine months before my birth, has never died. But by my metaphysical principle, if he has never died, he is still alive. Where, then, is Bob? But surely there is no mystery there. Every part of Bob—other than the cells in the placenta and the umbilical cord that were shed—developed continuously into a part of me, and every part of me has developed ultimately out of a part of Bob. It is thus quite futile to look for Bob outside of me. If Bob is anywhere, he is right here, where I am. It may be true that most of the original cells in Bob are no longer around, but that does not stop the survival of an organism: organisms replace their cells regularly and do not perish thereby.

Now, Bob can't be a mere part of my body, because all of my body has continuously come from Bob's body. Therefore, one can't set aside some special part of my body and say "that part of me is Bob." So, where is Bob? The answer is simple: Here. I am Bob. That embryo has grown to be a fetus, then to be a neonate, then an infant, then a child, then an adolescent and finally an adult. Bob is I and I am Bob. This was what I was trying to establish.

But this is a little too quick. I just said, vaguely, that Bob is here, and concluded that Bob is I. We need the following argument. Here where I stand there is only one large animal—Alexander Pruss. Bob is presumably right here—there is nowhere else for him to be. Bob has been growing for much of his life, and so Bob is also a large animal. The only large animal here is Alexander Pruss, and hence Bob and Alexander Pruss are one and the same animal. I, thus, am Bob. If Bob is here, and if no part of me is a large animal, and if Bob is a large animal, Bob and I must be one and the same entity.

Besides, given how organic development works, it is easy to see that every organ of mine is an organ of Bob's since Bob's organs have developed into being my organs, and yet without any transplant happening. Thus, I and Bob are organisms having all of our organs in common. But the only way that can be is if I and Bob are the same organism, i.e., I am Bob. "Bob"

and “Alex” are just different names for one and the same being: Alexander Robert Pruss.

There is only one way of countering this argument, and this is to deny that I am an animal, that I am an organism. This response seems absurd on the face of it, and it is right that we should see it as absurd. I am a rational animal. But there are three seemingly plausible ways of making this objection work. They are not the only ones, but they will be representative.

The first form that this objection can take is Cartesian dualism. Souls and bodies are separate substances. What I really am is a soul, a spiritual substance. The body is simply a tool that my soul owns and uses, much as I might use a hammer. My body is an organism, indeed an animal, but I am not myself an organism or animal. Thus, what Bob is is my body: an animal that I own. This dualistic view has various paradoxical consequences. My wife has never kissed me—she has only kissed Bob, my body. You cannot touch me—you can only touch Bob. Likewise, rape is then a mere property crime. Making philosophical sense of the meaning of sexuality is a lost cause: two persons’ having sexual intercourse is nothing but the intercourse between the animals owned by each of the persons. My body is simply my property, and so stealing one of my kidneys is a mere property crime—it is not stealing a part of me. These consequences are ethically unacceptable. After all, the government can morally take away some of my property for the greater good and does so in taxes. If my body were mere property, then the government would in principle have a right, when necessary, to extract a kidney from me as a tax payment. Finally, if this is right, then the traditional rallying cry of abortion supporters that “it’s my body” is no different in principle from the silly argument that I can do whatever I like in my house because my house is my property.

There is too much absurdity there, and so this Cartesian view fails.

But even if it did not fail, it could only be used by the proponent of abortion if he had good reason to deny that the soul sub-

stance was united with the embryo from conception—otherwise, the safer thing is to refrain from killing what might be I. But since the soul substance is unobservable, no such grounds are possible, apart from revelation-based religious arguments, and those should not be brought into a secular societal context.

The arguments against the Cartesian view are not arguments against the existence of a soul. The Cartesian view that the soul is a separate substance, distinct from the body, is not the only view of the soul. The Aristotelian or Thomistic view is that the soul is that which makes an organism to be the organism it is and to develop as it does. Thus, the soul is not something over and beyond the organism—it constitutes the organism as what it is, and what we are are organisms, organisms constituted by our souls. Thus, as soon as there is a unitary organism, there is a soul. (Admittedly, Aristotle and Thomas believed that the conceptus did not have the same kind of soul that I do—but they were theorizing in the absence of empirical evidence about the conceptus being an animal that continuously grows and develops into me, or else they were going against what they should have said by their own lights.) The Cartesian view is rather unpopular these days in secular circles. But there is a secular version of it, that replaces body-soul duality with body-brain duality: I am not my body and I am not an animal—I am a brain. This kind of a view will not help the abortion supporter all that much since the brain develops relatively early in pregnancy—around six weeks after conception. But in fact the most trenchant objections against the “I am a soul” view can be made against the “I am a brain” view. Only in the course of brain surgery can my wife kiss me if I am a brain. Rape, still, is only a property crime. My kidneys are not parts of me but mere property, and hence can be expropriated by the government if necessary.

And there is a further objection. My brain developed out of earlier cells guided by the genetic information already present in the embryo. There was, first, a neural tube, and earlier there were precursors to that. Brain development was gradual, cells

specializing more and more and arranging themselves. At which point did I come to exist? And why should the cells that were the precursors of the brain cells not be counted as having been the same organ as the brain, albeit in inchoate form? If so, then perhaps I was there from conception, even on this view.

The third response to my argument is that I am not my body or my brain, but what I am is my body's intellectual functioning. This response requires a metaphysical answer. On this view, I do not think. Rather, I am nothing else than thought itself, or more precisely, I am nothing else than a process of thinking. We would do well to reject this view just because it contradicts the commonsensical fact that we think. But we can also reject this view for a deeper reason. If I am a particular process of thought, then it follows that if that process of thought were not to have occurred, I would not have existed. Thus, when asleep, I do not exist. Moreover, were I not to have engaged in the processes of thought that I have engaged in over my lifetime, but instead were I to have engaged in different processes of thought, then I would not have existed—there would then have been a different process of thought, and hence someone else, if what I am is the process of thought that I am. It follows that we cannot think otherwise than we do because our very identity is defined by the process of thought we engage in. This fatalism, this deprivation of free will, is unacceptable.

As I said, there are views of who I am that compete with the view that I am an animal and that are not the same as these three, but they tend to be variations of these three. For instance, some think that what I am is a whole made up of two parts, a Cartesian soul and a body-animal. This view is open to the simple objection that two interacting parts do not automatically make for a whole. Moreover, there is the objection that surely I think, and yet my soul thinks, and since I am not a part of me, it follows absurdly that there are two thinkers here: I and my soul.

We see thus that I am Bob. I was once an embryo and a fetus. The embryo or fetus that was there was just I—in an

earlier stage of my life. This completes the first and hardest step of the argument.

An objection. In the first two weeks or so after conception, the blastocyst was not an individual, and hence in particular is not the same individual as I am, because it was capable of twinning—of splitting into two or more individuals—which in fact it does in about once every 260 cases. While what is normally called “abortion” is not likely to be done at this time since the woman at this time rarely knows herself to be pregnant, nonetheless there are abortifacients that act this early—for instance the IUD, Emergency Birth Control or the Pill in those cases where these act through an abortifacient effect—and hence the question is not merely of theoretical interest.

This objection rests on the false principle that if it is merely possible that an organism will split in the future, then we do not have a genuine individual on the scene. But this is plainly false: amoebae are certainly individuals, but they are capable of splitting. What happens to the individuality when they split is disputed by philosophers. One might hold that the old amoeba continues existing as one of the two new ones, but we simply do not know which one. Or one might hold, more plausibly, that the old perishes and a new one comes to be in its place. In the latter case, if I had had an identical twin, then I would have come to exist about two weeks after conception, not at conception, and the human being who came to exist at conception would no longer be alive.

But if we have an amoeba in front of us for a period of time during which it does not split, then it is the same amoeba, the same organism, over all of this time. This judgment is unaffected even should we learn that the amoeba could have split during this period of time, just as our judgment that someone is alive is unaffected by learning that she could have died yesterday. As long as the amoeba does not in fact split, it is one and the same individual as we had on the scene earlier.

One might argue that if one could know in the first two weeks that twinning was going to occur, then one would thereby

know that the conceived embryo would cease to exist at two weeks of age, and one could abort it earlier, since one would not be depriving it of a long and meaningful life. Whether this argument is correct or not—and I am inclined to think it is not, since I think how good the life that one is being deprived of should not affect whether it is wrong for someone deprive one of it—it does not matter in practice. We just cannot tell at the moment. And as in 259 out of 260 cases twinning will not occur, one needs to act on the presumption that it will not in fact occur.

## **2. IF I WAS A FETUS, IT WOULD HAVE BEEN WRONG TO KILL THAT FETUS**

There are several paths to the conclusion of the second part of the argument, that if I was once a fetus (or an embryo for that matter), then it would have been wrong to kill that fetus, under exactly the same circumstances under which it would be wrong to kill me now.

The most powerful argument is to look at what is wrong with killing me now. Killing me now is a paradigmatic crime-with-a-victim, the victim being me. What would make killing me now wrong is the harm it would do to me: it would deprive me, who am juridically innocent, of life, indeed of the rest of my life. Now, consider the hypothetical killing of the fetus that I once was. This killing would have exactly the same victim as killing me now would. Moreover, the harm inflicted on the victim would have been strictly greater, in the sense that any harm inflicted on me by killing me now would likewise have been inflicted on me by killing me when I was a child. I am now 29 years old. Suppose that left to nature's resources, I would die at 65. Then, killing me now would deprive me of years 29 through 65 of my life. However, killing me when I was a fetus would also deprive me of years 29 through 65 of my



life—as well as the years from the moment of the killing up to 29. Given that murder is a crime whose wrongness comes from the harm to the victim, it is clear that when the victim is the same, and the harm greater, killing is if anything more wrong.

Of course, there may be circumstances in which it is acceptable to kill me now. It might be that under some circumstances capital punishment is justified. If so, then it might be acceptable to kill the fetus under the same circumstances. However, it is also clear that the circumstances involved in capital punishment do not apply in the case of the fetus. Whether there are any other circumstances in which it would be acceptable to kill me now is a question that is beyond the scope of this paper, although I believe that the answer is basically negative.<sup>ii</sup> In any case, we see that the wrongfulness of killing me when I was a fetus is at least as great as the wrongfulness of killing me now in relevantly similar circumstances. Thus, my moral status when I was a fetus with respect to being killed is the same, or more favorable to me than, my status now.

The reason for the “more favorable than now” option is that we have an intuition that it is particularly wrong to kill people earlier. Although there may be no duty thus to sacrifice one’s life, we see nothing irrational in an older person sacrificing his life for a younger on the grounds that the older has literally less to lose by death. When I was a fetus, I had more to lose by death than I do now. Thus, to have killed me then would, strictly speaking, have been to inflict a greater harm.

Observe that nothing is said here about whether I was a person when I was a fetus. That issue is irrelevant. Whether I was a person then or not, killing me would have had the same victim and involved greater harm as killing me now. Observe that if I was not a person when I was a fetus, then the harm in killing me then would have been even greater than if I was a person then. For killing me when I was not a person would thus have deprived me of all of my personhood as lived out on earth, and this radical deprivation would have been a greater crime than killing me now which would not deprive me of ever

having had a personhood lived out on earth.

That said, an independent argument shows that in fact I was a person when I was a fetus. This gives a second argument for why killing a fetus is wrong, and it is the only argument I give that depends on issues of personhood. The argument turns on the metaphysical notion of an “essential property.” The essential property of a being is a property which that being cannot lack as long as that being exists. For instance, many philosophers think that being a horse is an essential property of a horse. If you take a horse like Silver Blaze and modify it to such a degree that it is no longer a horse, Silver Blaze will cease to exist and something else will come to exist in his place. Being material is an essential property of a rock: it could not exist without being material. Now, it is likewise plausible that being a person is an essential property of every person. If someone were a person and if personhood were removed from her, she would cease to exist. If this is correct, then the fetus that I was truly was a person since I am a person. If the fetus that I was were not a person, then it would be the case that I could have existed without being a person—which is impossible.

Even more plausibly, it is an essential property of me to have a property that I will call human dignity. Human dignity is a property of me that makes it wrong for another human being to set out to kill me when I am juridically innocent. As before, I leave capital punishment as an open question. Human dignity is an essential property: it is part of the essence of who I am. Were I to lack this intrinsic dignity, I would not be myself; I would not exist. But if human dignity understood in this way is an essential property and I have it, then the fetus that I was also had it—otherwise it wouldn’t be an essential property.

Finally, there is a very different argument for the wrongfulness of killing the fetus that I was, based on John Rawls’s concept of justice. Even though I take this concept to be incorrect, the more bases on which our argument can rest, the better the argument. Rawls bids us to imagine that we do not know which role in society we fill—imagining this is called entering

under the “veil of ignorance.” What kind of a society I would I come up with, and what kinds of rules would I rationally devise on selfish grounds, if I did not know which role in this society I am going to live in? Rawls says that that kind of society is the just society, and its rules are the rules of justice. In such a society, for instance, we would forbid racism because under the veil of ignorance we would not know whether we would end up having the role of victim or infliker of racism, and we would not want to take the risk of being on the victim. Likewise, we would prohibit the murder of adults.

Would we forbid the killing of fetuses? This question depends on just how much we are to be ignorant of under the veil of ignorance. If we know that we are not fetuses, then we might not forbid the killing of fetuses when it is convenient to non-fetuses because we would have no selfish reason to prohibit it. So, is the fact of us not being fetuses something that is under the veil of ignorance or not? Well, we must be careful not to take too much out from under the veil. For instance, if racism is to end up being deemed unjust, our race must lie under the veil. Moreover, even our being conscious must fall under the veil—thereby showing how much the veil is just a figure of speech since we cannot really be ignorant of our consciousness. The reason our being conscious must fall under the veil is that otherwise we might well enact that it is right to kill the unconscious for the sake of the conscious—to use the man in a coma for medical experiments, say. But at the same time, we cannot put too much under the veil. We had better have an awareness of ourselves as human since otherwise our “just society” will end up prohibiting all killing of animals, and this would make even most vegetarian farming wrong because of the moles and voles and other animals killed in the process of farming, as someone has once argued.

So, where do we draw the line? I would propose this simple criterion. Under the veil, we are aware of which social roles it would be logically possible for us to fill, but not aware which of those roles we do in fact fill. It would not be logically possible

for me to fill the role of a mole in the ground—I would not be myself then. So I know, even under the veil, that I am not a mole. However, it plainly is logically possible for me to fill the role of a fetus—it is possible because I did fill the role of a fetus once! Thus, whether I am a fetus or not is something that must fall under the veil of ignorance, and hence the killing of fetuses will end up being prohibited in exactly the same way as that of adults: we just wouldn't want to take the risk that we might end up being a fetus that is being killed. Hence, justice requires a prohibition on killing fetuses in exactly the way in which it requires a prohibition on killing adults.

Later, Rawls modified his criterion by talking of an unselfish caretaker for someone making the decision under a veil of ignorance about what role her charge would fill. This takes care of the problem that we can hardly be ignorant of whether we are conscious, while a caretaker can be ignorant of whether her charge is conscious. But it does not affect the rest of the argument. The caretaker needs to be ignorant of some properties of her charge, such as the charge's profession in life, but not of others, such as that her charge is not an insect. Again, I would suggest that a natural way to draw the line is apt to make the caretaker be ignorant of whether her charge is a fetus or not. For at the very least the caretaker should be ignorant as to which of the roles her charge could fill she in fact fills, and certainly her charge could fill the role of a fetus. And if that were so, the caretaker, truly loving whoever is entrusted into her care, would not want to take the risk of enacting a system whereby her charge could be killed.

### 3. IF IT WAS WRONG TO KILL ME WHEN I WAS A FETUS, IT WAS WRONG TO KILL ANYONE WHEN HE IS A FETUS

If you cut me, do I bleed any more than the next guy? No. I was not and am not special. If it was wrong to kill me when I was a fetus, it was likewise wrong to kill anyone else when he was a fetus.

It might be argued that there are some special differences between the fetus that I was, which we have seen it would have been wrong to kill me when I was a fetus, it was likewise wrong to kill anyone else when he was a fetus. It might be argued that there are some special differences between the fetus that I was, which we have seen it would have been wrong to kill, and some other fetuses. For instance, I was wanted. But that I was wanted did not anywhere enter into my arguments against killing me when I was a fetus. It is wrong to kill me now no matter whether I am wanted by others or not. Killing me earlier, I have argued, is not significantly different from killing me now, and so whether I was wanted or not is irrelevant.

A different objection would be that, as far as I know, I did not endanger my mother's life. However, my arguments would continue to apply even if I did: the fetus needs to be protected at least to the extent to which we would protect an adult under relevantly similar circumstances. If the fetus endangers the mother's life, it does so unintentionally. Whether it is acceptable to kill the fetus under those circumstances depends on whether it would be acceptable to kill me now were I to endanger my mother's life unintentionally. As I announced, the aim of this paper is limited: it is to argue that killing fetuses is wrong under the same circumstances under which it is wrong to kill adults, but it is not the paper of the paper to discuss the circumstances, if any, under which it is permissible to kill adults. I think it would not be acceptable to kill me were I en-

dangering my mother's life unintentionally: I will simply say in support of this that were I alone in a space capsule, three days from rescue, with my mother, with only enough air for 1.5 days each, it would not be acceptable for my mother or her agent to kill me.

A yet different objection is: I was a healthy fetus, but some others are not. The wrong in killing me when I was a fetus would have been depriving me of a meaningful and long future life. But what if the fetus cannot be expected to have such a life? Again, I respond that the purpose of this paper is limited: I am not going to settle issues of euthanasia here. It is acceptable to kill such a fetus only if it is acceptable to kill an adult who cannot be expected to have a meaningful and long future life. Again, I think it is not acceptable to kill an adult under such circumstances. Human life is intrinsically worthwhile and always meaningful. But this is not a paper about euthanasia. If I have shown that the fetus is worthy of at least the same respect as an adult in comparable circumstances, I have done my task.

# *Part 22: The Abortion Debate (Pro-Life)*

For this module, thus far, we have been looking at one of the best arguments which can be made on the pro-choice side. This argument goes after the core beliefs of the pro-life side. Namely, that if a fetus is a person, then abortion is murder (and thereby wrong) and that a fetus is in fact a person. The argument does not rely on consequentialist reasoning nor on a woman's bodily rights.

For the sake of fairness, I feel that it's appropriate for us to also look at one of the best arguments from the pro-life side. This is a purely optional reading as you can complete the assignments without reading it. However, it may serve you well if you ever encounter puzzles like this or if you are on the pro-life side of the debate, this paper will give you a far stronger argument for your stance. The paper is called *I Was Once a Fetus: That is Why Abortion is Wrong* by Alexander Pruss. His argument does not rely on religious beliefs nor on any notions of person-hood (per se). The argument goes like this:

- 1 I was once a fetus.
- 2 If I was once a fetus, it would have been wrong to kill that fetus.

- 3 If it was wrong to kill me when I was a fetus (that fetus), then it would be wrong to kill anyone while they were a fetus.
- 4 Therefore, it is wrong to kill anyone while they are a fetus.

The conclusion, the fourth line, makes all cases of abortion morally wrong. It may seem a bit strong, but that is what is argued. We will go through the paper point by point.

## I was once a fetus

This is the first line of his argument and it is the most seemingly obvious of the lines. Pruss starts off by asking us a few rhetorical questions (avoid them in philosophy papers, they just lead to confusion if they are not well crafted and obvious). The point of them is to show that at some point in the far past 'I' was an embryo, then a fetus, then a neonate, then an infant, then a toddler, a child, an adolescent, and then, finally, an adult. There are several ways in which a person can show that they were once a fetus. Pruss does this in an interesting way, one which I would not have thought to apply to this debate. he uses what we in the philosophy biz call a 'continuum', or an indiscrete series. To make these philosophically interesting, you need to have it such that something clearly holds on one side of the spectrum and clearly doesn't hold on the other. These are often called Sorites Paradoxes. 'Sorites' is Greek for 'heap' or 'pile'. These sort of cases are where we get the stereotypical philosophical question "how many grains of sand does it take to make a heap?" So, imagine that you have a pile of sand, a large pile, something which is clearly a heap of sand. If you take 1 singular grain of sand off of the top, is it still a heap? Clearly and obviously it is. Now, I remove another, and another, and another. All the while, 1 grain of sand doesn't make a difference, all the way down until I have only 1 grain of sand left. This grain of sand is clearly not a heap, but the reasoning shows that it is. For another example, take a spectrum of colors, from



blue to red. The far side is clearly blue and the other is clearly red. What if I start at the blue end and move over, ever so slightly, is it still blue? Yes, yes it is. Now, what if I move over a little more, still blue, a little more, still blue, and so on. In the middle, we will likely find something which we would call 'purple', but, by the reasoning, it would still be blue. As we continue, the responses get seemingly more and more absurd, until we are saying that red is blue.

Any line which we draw in this spectrum, saying that before this point, it's blue and after this point it's red or some other color would, frankly, be unnatural and arbitrary. In a third example, one which was given to me when I learned about this paper in the equivalent of 101 which I took (that course was exclusively epistemology and metaphysics, no ethics), concerns baldness. My professor, at the time, was balding, and was the example person for the entire department for this concept. On one side, you have him when he was young, full head of glorious hair. On the other side you have what he will become, a chrome-dome. Now, where is he now? Somewhere in the middle, is he bald? No, but is he 'haired'? No. There's the paradox, something is both bald and not bald.

Now, at this point, the same sort of reasoning can be used for the developing embryos through pre-birth infants all the way to you reading this in-front of some screen. You, now, are clearly you. You 1 second ago is also clearly you. You 2 seconds ago is also clearly you.<sup>165</sup> We can keep going back, and back, and back, deducting seconds off the clock, until we get to your birth. Before this point, it gets a little more tricky to imagine it. The continuous nature of the transitions are even more evident in this case. We have, now, that at your birth, that's you, and on the opposite side, at conception, we have a zygote, or some such. But this is again, a continuous spectrum. There's no hard-line, no non-arbitrary position where we could, within reason, say that you appeared and the zygote disappeared. Without such

---

<sup>165</sup>I am not sure about the grammar of these two sentences.

a line, then you are that zygote. So, this first line is pretty obviously true. I was once a fetus.

## **If I was a fetus, it would have been wrong to kill that fetus**

The last line of the argument showed that I was once a fetus. It is worth noting, and this will come up again later, that I could have used any one of you in place of myself. Now, our job is to show that if I was once a fetus (I was), then it would have been wrong to kill that fetus (me). We need to show that there's a causal connection between me (or you, or anyone) being a fetus and the wrongness of killing that fetus. There are two ways which I would go in trying to show this connection, if I were Pruss. Here's the route which he takes (along with my connections to previous content in this course) and then there's my personal preferred route. We will start by looking at the route which Pruss takes and then move on to the route which I would take. Both involve an intermediary proof for the stance that it's wrong to kill me now.

### **Pruss' way of showing that it's wrong to kill that fetus**

If you remember way back in the course, we covered Nagel's work *Death*. In that work, Nagel came from the idea that death is always bad for the person who died because something of great value is lost. That thing must be of such great value that losing it is worse than anything which can come from having it. For Nagel, this is experience. Having experience is so valuable that death is always worse than whatever experiences may come from the experiences. Connecting this into Pruss, my untimely death now would be cutting off experiences which I would have otherwise had. This means that killing me has deprived me of some good, and that makes it wrong. This is, interestingly, in

line with the typical Consequentialist kind of thinking (as in the way of approaching it is how the Consequentialist would do so). This is not Utilitarian thinking, however, because experience is included as an ultimate, super, good, which Utilitarianism does not incorporate.

So, we have that it's wrong to kill me now, which is awesome, but how do we move from this to killing me as a fetus? Well, we need to look at why it's wrong. As I have likely mentioned several times now, the why is the big aspect, this is the aspect which we can use in debate, either for or against, and make headway. To reiterate, it's wrong because I was deprived of experience. Suppose that I'm 28 and I would have otherwise lived to be 65. Killing me now would deprive me of 37 years of experience. Had I been terminated as a fetus, I would have been deprived 65 years of experience. From this, because 65 is greater than 37, we can say that killing me as a fetus is at least on par with killing me now, if not worse than killing me now. Because of this, we get that if I was a fetus, then it would be wrong to kill that fetus. Putting this as a more formal argument, we have the following:

1. Depriving someone of a good which they would have otherwise had is wrong.
2. Killing someone deprives them of a good which they would have otherwise had (namely experience).
3. If I was a fetus, then killing that fetus would have deprived me of a good which I would have otherwise had.
4. Therefore, if I was a fetus, then killing that fetus would have been wrong.

### **An Argument Which He Could Have Used**

Pruss' way of getting to the premise relies on the intuition which Nagel uses. This is not shared by many, but there are those who

have it. An alternative way which Pruss could have used to get this line, but does not (for reasons which we will see later), is to reapply the Sorites style thinking to this problem. All one needs to make this reasoning work is that it's wrong to kill me now and that the morality of killing another person does not change with time (which is Non-consequentialist style thinking). The first half, that it's wrong to kill me now, can be gotten through your preferred method, though non-consequentialist thinking will likely be preferred. To start, if it's wrong to kill me now, then it would have been wrong to kill me a second ago. This is because the morality of killing someone does not change with time. This also means that if it's wrong to kill me now, then it would have been wrong to kill me two seconds ago, and three seconds, four, and so on. This goes all the way back to when I was a fetus. This means that if I was once a fetus, then it would have been wrong to kill that fetus. Put as a more formal argument, we get the following:

1. It is wrong to kill me now (at T).
2. If it is wrong to kill me now (T), then it was wrong to kill me a second ago (T-1).
3. If something is wrong at T-n, then it is wrong at T-(n+1).
4. Therefore, If I was fetus, then it would have been wrong to kill that fetus.

There is a very innocent jump between lines 3 and 4 which can be glossed over. Basically, to make that move, there are some additional, veiled, premises which concern the time I was a fetus and the identity between myself and that fetus. The main point of this argument is that there's no hard line which makes it OK to kill me before a certain point, and wrong to kill me after that point.

**If it was wrong to kill me when I was a fetus, then it was wrong to kill anyone when they were a fetus**

The author’s original line used the gendered pronoun which I have replaced, but the point remains the same. Proving this line in a way that makes the entire thing still valid is a bit more difficult than normal. You see, this will appear to be inductive, which is not strong enough for philosophy, but it’s not. As I have mentioned a few times in this proof, I am not special. I could have put that you were once a fetus and ran the argument in the same way and I could have put any person in this argument in place of myself and gotten the same result. Since there are no cases where the validity of this argument does not transfer to another, we are safe in making the generalization to all people. Take these two arguments, which are oddly similar:

| I Argument                                                                          | Sally Argument                                                                                 |
|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| I was once a fetus.                                                                 | Sally was once a fetus.                                                                        |
| If I was once a fetus, then it would have been wrong to kill me when I was a fetus. | If Sally was once a fetus, then it would have been wrong to kill Sally when Sally was a fetus. |
| Therefore, it would have been wrong to kill me when I was a fetus.                  | Therefore, it would have been wrong to kill Sally when Sally was a fetus.                      |

The first argument is the one which we currently have a proof for. But, if we take the reasoning and replace the references to myself with references to Sally, we get the same result. Now, what if I was to replace the references to Sally with some generic, like, I don’t know, a human person. Take these for the comparison:

| Sally Argument                                                                                 | Human Person Argument                                                                                  |
|------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| Sally was once a fetus.                                                                        | A human person was once a fetus.                                                                       |
| If Sally was once a fetus, then it would have been wrong to kill Sally when Sally was a fetus. | If a human person was once a fetus, then it would have been wrong to kill them when they were a fetus. |
| Therefore, it would have been wrong to kill Sally when Sally was a fetus.                      | Therefore, it would have been wrong to kill a human person when they were a fetus.                     |

But, a human person is just a generic stand in, if this reasoning works (and it does, soundness is a different story), then the reasoning applies to all, real or potential, human persons. This means that I can move from the individual case, that it would have been wrong to kill me when I was a fetus, to the stance that it would have been wrong to kill anyone when they were a fetus, because the justification applies just as well to them as it does me. This gives us the final premise of the argument, namely, if it was wrong to kill me when I was a fetus, then it would have been wrong to kill anyone when they were a fetus. This last line connects everything together, and with some pretty easy reasoning, we have that killing anyone when they were a fetus is wrong, which makes abortion always wrong.

- 1 I was once a fetus.
- 2 If I was once a fetus, then it would have been wrong to kill that fetus.
- 3 If it would have been wrong to kill that fetus, then it would have been wrong to kill anyone when they were a fetus.
- 4 So, it would have been wrong to kill anyone when they were a fetus.
- 5 If it would have been wrong to kill anyone when they were a fetus, then abortion is wrong (this is implied from the

definition of an abortion).

6 Therefore, abortion is wrong.

With that, we have what is likely the best pro-life argument which you will find. Your typical pro-life argument which we saw earlier relied on an equivocation between 'human' and 'person'. This one, however, relies on sorites reasoning and, maybe, some intuition about experience. This could, maybe, also be used to make a hole in Warren's reasoning as well. But, that being said, there are some objections to this argument which need to be addressed. It should be noted that Pruss' replies rely on the Nagel intuition which we have seen previously. To reply to these objections without the value of experience being so much would require a different line of thinking than the one which Pruss uses.

## **But I was wanted!**

This is rather cold, I admit, but there may be one important difference between fetus-me and some other fetuses, namely that I was wanted. But, you have to notice that no-where in the arguments for the impermissibility of abortion were the claims that I was wanted. It solely relies on the fact that I am a person and that I was a fetus. There are some cases where the pregnancy and child-rearing will result in undue and heavy burdens on the mother. In these cases, it seems like abortion should be available (according to some). This is not to say that those burdens will always be present in that person's life, but rather that they would not be removed or be very difficult to remove if she went to term. This sort of reasoning falls in line with one of the popular pro-choice arguments which I have given for Warren, namely the ones which rely on the undue burdens on poor women.

## Pruss' Reply

It is worth noting that nowhere in the argument was there something about the want of a child, nor was there anything about the life of the child after birth. Even if I wasn't wanted, there's still a loss, namely the fact that I would have otherwise had experiences. A person is a person, no matter how small. Morality, often, will force us into situations which we don't want to do, which might not be in line with our self-interest.

## But I didn't endanger my mommy!

That was me making a joke of the objection a bit. The point of it is that, as far as I am aware, there were no complications with my birth/time in the womb which added risk to my mother's life, which may make a relevant difference between me and other fetuses. There are some cases where a person must choose between the life of the mother and the life of the child. There are other cases, even more extreme, where the process (not just child-birth) of going to term will kill both. Such as cases of unidentified ectopic pregnancies. These cases make up 10% the deaths during pregnancy.

## Pruss' Reply

But, you have to notice, that we have to chose between the life of one and another person from time to time, though it is not common. More over, in cases where a fetus is endangering the life of the mother, it does so unintentionally. In cases where it is acceptable to kill a person who is endangering another, it is intentional. So, since it is unintentional, it is not acceptable to kill the fetus. Also, if it comes to a choice between the mother and the child, if we were to apply Nagel's thoughts, we should choose the child. Namely because the experiences had by the mother up until that point plus those which will be had by the child is greater than the amount had by the mother if



she were to continue living. This, however, does not work in cases where the fetus is not able to develop and the process of development will only kill both. For example, extreme cases of ectopic pregnancies.

## **But I was a healthy fetus!**

This becomes a debate about euthanasia (the killing of a person painlessly when they have some extreme health problem). Killing me as a fetus would have deprived me of a long life, but killing an unhealthy fetus would not have deprived them of much. For example, Peter Singer, who I have used as examples before, is a rather extreme Utilitarian, both in his work as well as in his actions. His charity work and donation practices are top-tier. However, when Singer was directing a bioethics center in Australia (which is a universal healthcare country), he was contacted by doctors concerning their ethical dilemmas. In one case, he was contacted concerning issues in the Neonatal Intensive Care Units, ICUs for newborns. Many of these newborns had a range of very serious medical conditions which often lead to a very slow and painful death. If the newborns actually did survive, they would need multiple extremely expensive surgeries which would drain the resources from the collective system (it's even worse in a system without universal healthcare) and they would be very disabled in various ways. In the extreme cases of these conditions, the ones which the doctors were contacting Singer about, all people concerned, the doctors, nurses, and even the parents, believed that the babies should not survive. In these cases, all the newborns could do is experience pain. As a result, many infants were being left in the ICU untreated, with minor adjustments to alleviate the suffering where possible. This was an very painful experience for all those concerned, on an emotional level and took up resources which would have otherwise been provided to healthy newborns. Almost all cases like these die before they are 6-months old. Peter Singer advised

that the newborns with the extreme cases of this condition be given a quick and painless death. Would it not be better for the parents to have this condition identified early and be able to have an abortion?

### **Pruss' Reply**

This becomes a debate about euthanasia (the killing of a person painlessly when they have some extreme health problem). The connection is that, for the author, killing an unhealthy fetus is the same as killing a person with a terminal illness. So, if you think that killing a person with a terminal illness is permissible, then you must also think that killing an unhealthy fetus is permissible. The author thinks that euthanasia is wrong, full stop. Again, this is a lot like Nagel's reasoning. The core principle is that no experience can make life not worth it. Because of this, every human life, for Pruss, must be worthwhile.

### **A Reply to this Reply**

There is an issue with this reply. Namely, if experience is what makes human life always worthwhile, what if the fetus is going to be brain-dead? For example, what if the child has a certain genetic abnormality or some other factor which results in anencephaly (Greek for no-in-head (AKA brain)). This is a real thing, the baby, if it has the main brain stem, will be alive, but will not have any experiences whatsoever. No consciousness at all. The life expectancy for a baby born with this is between a few hours and a few days. This means that the human's life would not have anything to it, there would be none of what makes life always worth it, according to Pruss. This means that he does not have a way of saying that abortion would be wrong in this case.

## MODULE X

*What about duties to  
animals and future  
generations?*

## *Part 23: Vegetarianism/Veganism*

Vegetarianism, in a nutshell, is the stance or practice of abstaining from the consumption of animal products. This is an umbrella term which encompasses several different variations on the stance. For example, you have ovo-lacto-vegetarianism, which says that eating eggs and milk is fine, but the flesh of the animal is off limits; there's lacto-vegetarianism, which says that milk and dairy products are fine, but eggs and flesh are a no go; ovo-vegetarianism says that eggs are cool, but milk and flesh are off-limits. There are several other varieties aside, for pretty much any kind of animal product which you can think of, there's likely a flavor of vegetarian which says that it's off limits. Because of the kind of creatures we are, we can't totally cut out all foods which come from a living thing (this would say that plants are off limits too), so the overarching theme for vegetarianism is that the food consumed and (maybe) the products worn/used must be predominately plant based. The most extreme version of vegetarianism is veganism. This is the stance that all animal products, and even insect products, like honey, are not to be eaten. With the stance clearly laid out, we need to discuss the why, remember the why is always the most important part.

## Why are People Vegetarian?

Some people out there, like some of my mother's side of the family, think that vegetarianism is a new-fangled thing, something 'hippy-dippy'. But, this idea and the practices which surround it are very old. Many ancient Greeks practiced it and many religions and cultures have vegetarianism built in because of the moral implications. Plato believed that the consumption of meat did not belong in a civil society because it caused violence in the soul. Pythagoras of Samos, who you may remember from a math course, and his disciples refused to even talk to hunters or butchers. Some very old customs in Asia have vegetarianism built in on religious grounds (also, by extension moral grounds). For similar reasons as Plato and other Greeks, some early Christian sects practiced vegetarianism. Leonardo da Vinci, because of his concern for animal welfare, abstained from the consumption of animal flesh. Mary Shelley, the author of the world-renowned book *Frankenstein*, because of the deep appreciation for the beauty in nature which came from the art movements of her time, refused to take part in the killing of an animal, and thereby refused to eat meat. And then, also, you have me. I have been vegetarian for 20+ years, this was initially because of concern for animal welfare, the realization that what I was eating once was a thing which could feel pain, but this evolved over time. I started looking into my family history and a solid chunk of my family are/were vegetarian and they did not have various health problems which tend to appear in my family (diabetes, mostly). But, I am not saying that this is healthy for everyone, just how it is for my family. Abstaining from animal products for health reasons is a relatively new concept and there is some evidence to back up the claims that limiting the consumption of meat is good for you, but it's far from common. Looking at the examples I just gave and the current trends, we see that people are vegetarian for moral reasons. But, what are those reasons?

With the exception of some stances which may or may not

entail a vegetarian diet, by far the most common reason for people to adopt this stance comes from some moral intuition. In this case, the intuition comes from a concern for animal welfare. But (and this might seem obvious), what makes eating meat violating to animal welfare? Put into ordinary terms, this stems from the idea that animals suffer, they feel pain. The most appropriate moral stance to take on this sort of consideration is consequentialism. Consequentialism says that causing more pain than necessary for some greater good is wrong. The next aspect comes from the realization that the process of killing and butchering animals for food does cause more suffering than necessary for some greater good. So, the process of killing and eating animals for food is morally wrong. Since Consequentialism does not see a distinction between killing and letting die (I don't), eating the animals is a cause for their death, so you are, in a sense, causing this suffering, which the reduced consumption would have prevented. So, eating meat is wrong.

### **The Animal Suffering Argument**

- 1 Animals feel pain.
- 2 Causing pain when it lacks a pleasure which outweighs it is always morally wrong.
- 3 Killing an animal for food causes more pain than pleasure (killing animals for food is a case of (2))
- 4 Therefore, killing animals for food is morally wrong.

It should be noted that the first line of the argument is not necessary, it's sort of built in to the third line, but it makes it clearer for what Descartes is going to be replying to. This argument serves as a good backdrop for discussing whether or not we should all be vegetarian/vegan. Morally speaking, we are in a situation where, though we come from meat eating ancestors, we don't need to do that anymore to survive, so it might be time to remove it. As is always the case, if a person makes a claim, they need to back it up with reasons. For example, if some-

one was to claim that Peter Quill dating Gamora<sup>166</sup> is morally wrong, we would ask for some reasons. If someone claimed that the earth is flat, we would ask for proof, reasons, same as in the case of it being round. The Animal Suffering Argument is a set of reasons for us to be vegetarian/vegan and the reasons are structured in a valid way. Since the argument is valid, we need to look at the facts of the case. Now, what we need to do is see what objections the opposing side may have (but if they want to say that everyone should not be vegan, then they need to make an argument for this).

## Descartes' Reply

Descartes, who you may remember from this class, had an interesting reply to this argument. As I mentioned in the previous section, this line was not necessary but it's mostly there to help you see the reply which Descartes is making. Basically, Descartes claims that animals don't feel pain. Without the first line, Descartes would be presented as claiming that killing and eating animals for food does not cause unnecessary suffering; not because we 'need it to survive', but rather because it does not cause suffering.

## The Mind-Body Problem and Ethics

Remember that Descartes was our case of a substance dualist. He held that there were two kinds of things in the world, minds and bodies. Suffering, pain, emotions in general, are all mental, not physical (though the reason we think others have these is because of the physical responses given). If you are a physicalist, of any kind, then you might not find this persuasive.<sup>167</sup> From substance dualism, we get that suffering or

---

<sup>166</sup>Guardians of the Galaxy part 1

<sup>167</sup>It is possible, however, to be a physicalist but claim that only the human brain or greater is complex enough for consciousness to emerge or some such like that.

any other kind of experience, positive or negative, is something which requires a mind/soul to have. If Descartes were to say that animals suffer/feel pain, then he would need to say that they had souls/mental substances. This was a step too far for Descartes. Descartes thought that humans had souls, but animals did not. Going outside of what Descartes thought, we could get reason to think this as a reply to a particular version of the Problem of Natural Evil (concerning evolution), though Descartes might not have given this thought:

- 1 The process of evolution is necessary to get beings such as us with free will and moral aspects.
- 2 The process of evolution, if animals have souls, would result in an incalculable amount of pain and suffering.
- 3 An all good god would not allow for this sort of thing.
- 4 There is an all-good god who would want beings with free will.
- 5 So, animals don't have souls.

When I was in undergrad, there was a seminar which I took on the problem of evil as it relates to evolution, this was an argument which appeared there (put in my own way with my own thoughts) Since Descartes has his dualism, we can get that since animals don't have souls, they don't suffer. To make this point clear, Descartes did not argue this way and it was an argument I made up for this content, nothing more.

An injured animal, for Descartes, is nothing more than a broken machine. If animals do not feel pain, then there would be no suffering caused by the killing and eating of them. Poof, all moral reasons for not eating meat go away (when it comes to the animals themselves).

## Some Replies to this Objection

Here I am going to give a few of the replies which people could give to Descartes' idea that animals don't feel pain. These come



in three different flavors, each of which we have seen in various modules throughout this course, not necessarily the Mind-Body Problem.

### **Physicalism:**

Remember what was said about Physicalism. This was in the Mind-Body Problem. This is the stance, roughly, that there aren't things like souls. This puts a hole into Descartes' reply to the Animal Suffering Argument. The physicalist agrees with Descartes that animals don't have these mental substances, but neither do humans. They claim that there are no mental substances, only physical ones. This points a whole in the core assumption in Descartes' argument. Whatever way the physicalist has to explain mental phenomena, such as pain, will likely apply equally well to sufficiently complex animals. One of those is clearly flawed.

### **Related Objection:**

Following this line of thinking, concerning the Mind-Body Problem, if we look really closely at the dualist stance, we see that there's nothing built into it which entails that non-human animals don't have souls and that human persons are the only ones with them. Descartes snuck this assumption into the stance and there's no reason to think that it's correct. There are, potential, arguments for it, bringing in things like the Free Will Debate or other assumptions, but these are equally warranted to pose objections to them. Similarly, there's nothing in dualism which states that souls have these properties and there's only one kind of soul. There could be a series of different kinds of souls with different experiential capabilities, much like there are different brains with different levels of complexity. Humans, as far as we know, just so happen to have the highest quality soul, so to speak, and even some humans lack those.

## Darwinism:

This objection comes in part from the arguments for and against the existence of God. Darwinism, evolution, gave us a way to explain the complex structures in the world without the need for a divine architect. But, it also gives us a way of working backwards to get that animals feel pain. This does not necessarily require physicalism, but that stance fits nicer into the idea here. Creatures on this planet evolved from a common ancestor which had certain features, these features remained because it was adventitious. Having the capacity to feel pain would be adventitious for creatures to have and the beings which evolved from the common ancestor with that trait would have retained it. So, at least some animals feel pain, in virtue of the relevant features they share with humans, like certain kinds of nerve endings and brain structures.

It could be said that a soul evolved in conjunction with these complex brains necessary for pain, which would mean that they came in degrees, or it could be an addition to a physicalist model. Either way, we get this.

## Utilitarianism:

This particular reply is mostly here to point out how unintuitive the idea which Descartes is proposing is. Objections like this, in Philosophy, are called the "incredulous stare." These are cases where the stance is internally consistent, we can reason and debate through it, but it's so wild and out-there that we just don't know how to reply aside from staring at them in disbelief. When we think about the rightness and wrongness of our actions, often, the pain and suffering we expect to cause fits into the reasoning. For example, Jeremy Bentham once wrote "The question is not, Can they reason?, nor Can they talk? but, Can they suffer?"<sup>168</sup> Though it may be true that they can't suffer to the degree which we can, it seems that they

---

<sup>168</sup>Bentham.

can still suffer to some degree. In many people's calculations of the consequences of actions, animal suffering does play a role. Though it may be seen as more basic (bodily) than the ones humans can experience, they still count for something. As a result, a sufficient amount of suffering could outweigh the bodily pleasure which the fried bacon gives.

## To Get Out of These Worries

If you wish to claim, as Descartes does, that animals don't feel pain and avoid these worries, you will need to do a few things. First, you will need to reject the idea that species can change over time (also called evolution), and in doing that (to stay consistent), you will need to reject the idea that things change over time, which is easier for some than others. Second, you will need to say that all people who think about animal suffering in their moral reasoning are wrong, which is harder for everyone. You can do this by either accepting Kantianism (which has another way of getting at the animal suffering argument later) or by claiming that animals aren't worth adding to the moral figuring (either way, me kicking an unowned dog is morally OK). And third, you will need to be a dualist, but one which rejects the idea that animals have souls and, because of that, you will need to claim that there is no doggie/kitty heaven.

## Speciesism and Anti-Speciesism

Speciesism is another way to get out of the argument, basically meaning that you can have as much meat as you want and not be a moral monster. Speciesism is a fancy term and it should remind you of terms like 'sexism' and 'racism'. It has a very similar origin in how we use it. For example, racism, roughly, is treating one group differently than another group solely in virtue of their race (assuming all else is equal). For example, it would be racist to put a Mexican person in jail for jay-walking

for 6 months (and that's the only offense) while only putting a white person in jail for the same crime for 1 month. Similarly, sexism, roughly, is treating one group differently than another solely in virtue of their sex. For example, (assuming that all qualifications are equal) it would be sexist to give a man a promotion over a woman, solely because he is a man.<sup>169</sup> Speciesism follows a similar line of thought. Speciesism, roughly, is treating one group differently solely in virtue of their species. For example, there used to be a Comedy Central show called *Ugly Americans* in which the core premise is that the various monsters (zombies, vampires, demons, etc) are real and are trying to integrate into New York. Refusing to offer a job to a demon solely because they are a demon would be speciesist. Almost any Science Fiction or Fantasy story where there are more than one intelligent species could be used to make examples, all following the same line of reasoning as racism and sexism. But, speciesism does not just apply to these strange, unreal, stories, but it also applies to the real world as well. Speciesism would tell us to favor the interests of a human child over an ape, solely because they are human, no other reason is necessary. Though there may be other reasons, those reasons are extra. There are some reasons to be a speciesist, and many people think that reasons like these are persuasive:

## Arguments for Speciesism

For this, if these arguments are persuasive, then you are being speciesist. Applying this to the Animal Suffering Argument, they claim that the suffering of an animal only counts morally when humans are not involved or aren't benefited. If the animal suffering benefits humans, then the benefit to humans is all that matters.

---

<sup>169</sup>These examples also apply in reverse, it would be sexist to give a woman that promotion merely on the grounds that she is a woman. We would need to find other qualifying factors which should be taken into account, otherwise flip a coin.

## Christianity/Abrahamic Stances:

This does not just apply in Christianity, but also in most religions which have a similar historical origin (Islam, Judeism, and others). In such belief systems, you have a passage or belief that God created man and gave them dominion over the world, as in God created humans with the explicit feature that they are greater than all other beings, and as a result all humans are worth more than animals. Morally speaking, no amount of animal suffering (if they actually have it) counteracts human pleasure. But this stance does have its issues, as we will see.

- 1 God gave man dominion over the animals.
- 2 If (1), then they aren't equal.
- 3 If they aren't equal, then animal suffering, if beneficial to humans, doesn't count morally.
- 4 Therefore, animal suffering, if beneficial to humans, doesn't count morally.

This argument, however, does have its issues. For example, this assumes the existence of God, which is debatable, but it also assumes that dominion means that we can use them as we see fit without moral consideration. This might not be the case. For example, just because I own a dog, I have dominion over them, but that doesn't mean that I can treat the puppy poorly.

## Tu Quoque:

This is a Latin term for "you also". Think about it this way, imagine that a parent is a smoker and gets really angry when they catch their 21+ year old child smoking; a natural response is for the person to say "well, you smoke, why can't I?" Basically, if you do it, I can too. This is a fallacy, just because another being does something, it doesn't make it OK. In the animal kingdom, we don't see other animals treating other species without favoritism towards their own (as in, we see animal favoring members of their own species over members of different

ones). This raises the question "why shouldn't we favor our own, solely because they are our own?"

The issue with this line of thinking is that it's a fallacy, the core root of the reasoning doesn't generalize well. For example, (this is an example given to me by former military students) in some cases, the enemy combatants in war will not treat our fallen soldiers with respect. The question here is "are we allowed to treat their fallen disrespectfully?" Unless you are particularly hardened, the reply is "no, we are better than that." And that's the point, we are better than that, we need to hold ourselves to a higher moral standard.

### **The Children Analogy:**

This is an example that I tend to give as an example in Feminist Ethics, but it works quite well here. There, it's used to show that bias can be a morally good thing and being impartial wrong (Feminist Ethics tends to have a level of bias built in as a good thing). This example tends to work best for people who are/were the primary care giver to children (the mom/Mr. Mom).

Suppose that your child and their friend are swimming in a lake, due to the cold water (common in Washington), they both begin to have difficulty breathing as they swim, resulting in them starting to drown. You can only save one, who do you save?

For people who raised children, this tends to be a very easy answer "I save my kid." Feminist Ethics uses examples like these to argue against that impartiality is a feature of correct moral thinking, but here, we can use it to work in another way. People tend to favor their own children when their interests conflict with others. We don't see this as morally wrong. A very similar case can be seen in favoring members of your species, why is that seen as wrong? For example, let's look at this case, the

### Chimp or Baby Case:

Suppose that a fire has broken out in an apartment building. Everyone has gotten out with the exception of a small child and a chimp. We will suppose that the child has certain disabilities which makes them the mental equivalent of the chimp. Yet again, you can only save one. Is it wrong for you to save the chimp?

In this case, if you say that it is wrong to save the chimp, then you are being speciesist.

### Anti-Speciesism

Anti-speciesism, on the other hand, does not imply that when choosing between a dog and a baby to save, you should flip a coin, but rather you should make your choice based on factors other than whether it is a member of your species, like the capacity for pain. When it comes to moral considerations, the species of the creature does not play a part. This is central to Consequentialist style ethical thinking. You can also think of it as that the painful experience of one creature should be weighed the same as the same experience had by another creature. Humans are not special in the grand scheme. Rather, we are only special in that, as far as we are aware, we are the only beings which can experience certain kinds of mental suffering. So, in the Chimp or Baby Case, the anti-speciesist would say that the two lives are equal, but in the case of choosing between a worm and a human baby, you would choose the human baby, because that creature is the kind of thing which has a greater ability to experience happiness and contribute goods. It is certainly possible, however, to have an ample-enough number of creatures to outweigh the value of a human, for example, it might be possible that 5 chimps equal 1 human, or some other conversion system, depending on the context of the case.

## Animal Rights

Here we are going to be talking about rights. But, rather than talking about human rights, which is an entirely different course I teach, we will be looking at whether or not there's such a thing as 'animal rights'. Now, I don't know whether or not this is just a story or it actually happened; but regardless it makes a good story: Wesley Hohfeld tried to make law students carefully figure out the different ways that the word 'right' was used in American law. This made a lot of people angry and his students even tried to get him to lose his job as Chairman.

The lesson from this is that, since law students don't like the analysis of the various kinds of rights which we can have (passive, positive, legal, moral, etc), we should not be shocked that similar break downs are not liked by politicians, writers, and political theorists. Some have a vested interest in not using the term 'right' in various contexts, because it carries a certain weight, while others have a vested interest in applying the term in cases where it does not apply, for the same reasons. Some politicians can even have a vested interest in keeping that kind of talk (about rights) as meaningless as possible.

'Rights' is a term which gets thrown around a lot and often in inconsistent ways (for example the 'right to own a gun' does not seem like the same sort of thing as the 'right to freedom of expression'). But, for our discussions here, we will use this basic condition, one which is seemingly universal:

If something has a right to X (either do something or have something), then others have a moral duty to ensure that they have X.

This basically says that if you have a right, then others have the moral duty to make sure that it's not infringed upon. For example, if you have the right to vote, then I and all other people have the moral duty to make sure that you can vote (it's morally wrong for us to let your right be taken away). Similarly,



if you have the right to life, all other people have the moral duty to make sure that you don't get killed. If you have the right to a fair trial, then other people have the duty to make sure that you get it if needed.

But all of these examples apply exclusively to people, beings of a certain intelligence and maturity level. Some people want to extend the notion of rights beyond people and to non-human animals. Such people are often found in animal rights movements. For example, some might claim that animals, like people, have a right to life. This would mean that all others have a moral duty to ensure that they don't get killed or at the very least they have a moral duty to not kill them. Others might claim that animals have a right to a certain standard of living (like people, but this one is not commonly talked about). This would mean that people have the moral duty to not destroy their habitat. Other examples can be made and the extent of the moral duty would depend on the actual range of the right, same with that of human rights.

## **Some criticisms of animal rights**

There have always been, to some degree or another, in various cultures around the world, people who think that animals have at least some rights. They could argue that rights come in levels. If a being has certain features, then it has rights and the rights of those beings below it on the level. Human persons would be at the top of the chart, so to speak, so we get all of the rights, But, say, chimps and dolphins get some but not the right to vote or some such. This would be much like the features of person-hood found in the Abortion Debate Module. That being said, there are some criticisms directed towards the very notion of Animal Rights which would apply regardless of the system for determining which rights would otherwise be had.

## Rights Imply Responsibilities:

If animals have rights, then I have a moral duty to see that they have whatever the right is. This comes from the definition of rights which we are using for this section. But, at the same point, when we look at the various uses of the term 'rights', we see that it also implies something else. In this case, having rights gives us responsibilities. For example, if you have the right to own a gun, then you have the responsibility to use it properly and in the right kind of cases or if you have the right to vote, then you have the responsibility to know the details about what you are voting on (if you do so). If you can't handle the responsibilities which come with a right, then you don't have that right. We can't say animals have rights because they lack the responsibilities, or the ability to fulfill those responsibilities, which come with those rights.

## A Reply

To this idea, many people will point to how some humans lack the ability to take on the responsibilities of having various rights and yet still have them. For example, some claim that people in a permanent vegetative state have the right to life, but they lack the ability to take on the responsibilities of life itself. Children who cannot develop past a certain point due to genetic diseases or some such are still said to have rights despite not having the moral ability to have moral duties. We even have various systems in place which determine whether a human has certain rights or not depending on their intellectual abilities. Animals are no different.

## Direct Vs Indirect Duties to Animals

As I mentioned in the onset of the last little section, when a thing has rights, we have a duty to ensure that they get or have the ability to do whatever the right is. Immanuel Kant,

who you know from Kantianism, held that animals do not have rights, as they are not people, which means that we don't have direct duties to them, but rather we have indirect duties to them. When a being has a right to something, others have the duty to them directly to ensure that they get what ever the right implies. But, not all duties are direct, sometimes we have duties because of other duties. For example, (though Kant may not like this), I have the duty to feed my children, which means that I also have the duty to make money and buy them food. My duties there are indirect because I have them to complete other duties. My duty not to lie to you is a direct duty because it's one step removed from a person, you. My duties to (certain) animals are indirect because they are (at least) two or more steps removed from a person. My obligation to not kick your dog doesn't come from the animal, but rather it comes from you. I have a duty to you and that duty entails a duty to your animal.

It is wrong to harm certain animals because we have a duty to ourselves or other humans to not harm them.

For Kant, needlessly harming animals is wrong, not because of the pain, but because of how this is damaging to the character of the person doing it which we have a duty to ourselves/others to not harm. Maintaining certain character traits is a duty we have to ourselves because it leads to us better following the categorical imperative. According to Kant and a lot of popular culture, a person who beats/tortures animals is more likely to beat/torture humans, which would be using them as mere-means. Maintaining your character as an upright person prevents you from falling short on your duties to other people.

## Some Replies to Indirect Duties

This notion of indirect duties, or that we have duties at all, seems to run into criticisms here and there. For example, when

I give students certain ethical problems concerning the environment, they want to say that we don't have duties (meaning that it's morally wrong for me to fail to do the thing) at all, saying that doing it is going above and beyond. But there are other, more common criticisms which can be put against Kant's ideas here.

### **Implausibility:**

If you in the Ethics module did not find Kantianism particularly persuasive or you thought that it was loony-tunes and found the Consequentialist line of thought more appealing, then you will need to handle the Animal Suffering Argument in some way, but you don't need to think that we have duties to animals. Consequentialist thinking does not have room for the ideas and reasoning which Kant gives.

For the Utilitarian and other Consequentialists, the idea that morality rests on duties, indirect or otherwise, is just missing the point.

This is because they think that morality rests on the consequences of the action, not the character of the action or some other thing about human autonomy or some such. Morally speaking, rights are there because of the consequences of people having them, they don't come pre-built in, which leads to animals potentially getting them too.

### **Untested and Unconfirmed Empirical Claim:**

You have likely heard of cases where particularly heinous serial killers started with animals first. This might make you think that the disposition to harm animals is a slippery-slope to harming people. But, this is, frankly, an empirical claim, it's the sort of thing which scientists would need to perform tests and confirm. As far as I am aware, there is no data which links the two. But this does link into an interesting discussion of a certain kind of fallacy. Lottery groups frequently fill the

airwaves with pictures and stories of people winning the lottery (we will see this sort of fallacy again when we talk about political philosophy). These grab our attention and make us think that the winners are in, some way, close to us. This makes us think that winning the lottery is very common. But, in fact, it's quite rare.

Similarly, when it comes to serial killers starting with animals, we all know stories of this happening and it makes us think that it's common for people who harm animals to become serial killers, but it is be quite rare.<sup>1</sup> Does a willingness to treat animals as mere-means lead to a willingness to treat humans as a mere-means? Are farmers who beat their horses in drawing their carts more likely to beat their wives/husbands?

# *Part 24: What Are Our Duties to Future Generations?*

When I say ‘future people’, I mean people who aren’t born yet, who, unless something terrible happens, will exist a few generations down the road. People who aren’t even a twinkle in their pappy’s eyes yet. Although I can’t be sure that any one of them will exist, I can be sure that some will exist. One way to think about this interesting point is that there, more than likely, will be people in the future, but I can’t say with certainty who those people will be, what individuals will make up the collection of people out there. We can point to things like chaos theory<sup>1</sup> to explain why we can’t know who those people will be. Now, our question is ”do those people, people who don’t exist, but likely will, have rights? Are they the sort of things worthy of our moral consideration? Do we need to be concerned about their welfare?”

Some of you, I am willing to wager, will say that this is a no-duh sort of question, but, like with the Abortion Debate, we need to look at the reasons. Just as in that case, both sides

will say that it's obvious. Some say, the consequentialists being the strongest voices there, that we have the same obligations to future people as we do currently existing people. Just because they don't exist, they will, so we need to ensure that the best future is there for them. Others will say that we don't have any obligations to future people. Here we could find the non-consequentialists. They would say that, for example, the being must exist in order to be treated as a mere-means; so we don't have obligations to them. We can only have duties to contemporaneous persons. But this debate and difference leads us to this interesting topic:

## Intergenerational Justice/Ethics

Trying to figure out the morality of actions concerning future people is an area of philosophy called 'intergenerational ethics'. When it comes to their rights and the duties we have to them, this is 'intergenerational justice'. Since the consequentialist is not too much of a fan of rights and doesn't, fundamentally, think about morality in terms of duties, the consequentialist will tend to work in the more general intergenerational ethics, while the non-consequentialist will tend to work in the more particular intergenerational justice. However, there are certain powers which the current generation has over the future generations which are not had by them over us. These powers add variables into the typical equations which we would use for the morality of actions which lead to both epistemic and metaphysical questions which need to be addressed in order to figure out what the right course of action is. There are three such powers, with the third being the greatest and most perplexing problem, especially for the consequentialist.

### Limiting choices:

The current generation can set up a system which would be very costly for the future generations to change and, essentially,

force them to continue with that system. For example, what if we made various choices which placed the future generations into an extreme economic debt, this would force them to maintain certain policies in order to pay off said debt. Or, on a smaller scale, what if we bought houses which were tied to familial wealth with a 200 year long repayment plan? This would force the future generations to live in that house, unless they got very wealthy rather quickly. Future generations are not able to place that sort of burden on us, they can't force us to do anything (with the exception of if time-travel happens, then they could). But the limitations don't need to just be economic, they can also, for example, be intellectual. In the case of intellectual limitations, we can greatly advance in one direction which would make the alternatives woefully under-researched, making switching to the alternatives very hard because it would require the future generation to back-track and start the research from scratch. Here are two examples:

| The Green Dictators                                                                                                                                                                                                                                                                                                                                                                  | The Cheapest Route                                                                                                                                                                                                                                                                                                                                                                             |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Suppose that we know that the future generation will need to continue pursuing alternative energy sources and not transition back to coal/oil power for electricity/transportation. As a result, the current generation signs multiple treaties which make it very difficult for the countries to back out and have extreme penalties for refusing to comply with this green agenda. | Suppose that we build an infrastructure, roads, electrical lines, waterways, etc. with the easiest materials and power-sources available to progress very quickly and have great advancements. This infrastructure gets so ingrained that the transition to other sources of power and other infrastructure methods is very expensive and underdeveloped compared to where it would have been. |

For both of these cases, we have constrained the future generations to take a path which they did not choose for themselves. They could not have given consent or a voice in the process. So, we could, maybe, say that they were coerced into the system. But, some might say that we did the right thing



when it comes to The Green Dictators and the wrong thing in The Cheapest Route. This difference must be because of the consequences. But, we are still forcing another group to follow our choices, much like the Cultural Imperialist. As an interesting side note, regardless of the path we choose, we are still imposing our wishes onto the future generations.

### **Unidirectional Benefit:**

This is a sort of reiteration of the 'Limiting choices' ability which our generation has on the future ones. It is possible for the current generation to benefit themselves at the expense of the future generation and the future generation will experience all of the cost and, in some cases, none of the benefits. In such cases, the current generation will enact policies or engage in behaviors which will benefit them in the short term, be costly in the long term, and be long dead by the time those bills come due. The Cheapest Route case works for this example if we add in that the benefits of the rapid expansion are less than the costs to the future generations. But there are other cases:

### The Origins of the Automobile (this is sort-of fictional)

---

When automobiles (cars) were first becoming a product which the average American could buy, there were three different generic kinds. First, they had the steam-powered car. There were several companies producing these and they had the benefits of being familiar technology as well as able to go long distances. They weren't bought up as much because they didn't have the 'get up and go' and the speed of the other options. Second, there were the electric cars. Women consumers really liked these because they were just a push-button to get started, but they did not have the speed nor could they go long distances. And finally, there were the gasoline cars. These, with the introduction of the electric starter, had the quick start, the speed, and the ability to go long distances. Men, who were the predominate buyers, really liked these. That generation collectively mostly going with gasoline cars greatly benefited them, but it resulted in great costs to the future generations, in the form of Global Warming.

### Multi-generational Mortgages

---

In some areas of the world, there are multi-generational mortgages 100+ year plans (the most famous are Japan's 100 year mortgages, but Sweden has 105 year plans). In such cases, the future generations will be forced to pay the debt off of the house or the loan with the current generation getting the money from the loan. This will also force those generations to live in those houses until the house is paid off and able to be sold, some cases they are forced to sell the home to pay it off.

As I said previously, there are three, general, powers which the current generation has over the future generations which they don't have towards us. But this third problem deserves a page all to itself, because it leads us into a very complex and mind-bending problem, especially for certain kinds of consequentialists.

## Power to bring them into being:

Not only do we have the power, without any resistance from them, to essentially make them our slaves long after we have taken to dust, but we have the power to actually create them. We have the power to influence and alter what individuals will come into being and how many of them will come into being. We could, though unlikely, completely destroy ourselves, making no future people possible, or we could have a giant baby-boom which makes a ton of them. Even little things, which are totally coincidental or seemingly unrelated can result in an entirely different crop of future people being produced than if it hadn't happened. Take these three cases:

### The Black-Out Baby Boom

In certain areas in New York, there was a time when the places would shut down because everyone was watching Friends. On one such day, a young grad student and his wife were settling down to watch and there was a power outage, so, they grab a few blankets, light the candles. . . 9 months later, a baby is born. This was, to be exact, the August 14th Northeast Blackout in 2003.

### The Iron Maiden Baby Boom

In Brazil and several South American Countries, when Iron Maiden did their world tour, on Flight 666, there was a spike in births in the cities which they visited 9 months later, in order of their visit. (This is a bit exaggerated, but Woodstock had a similar effect.)

### The Sport Event Baby Boom

In 2005, the Red Sox beat the Cardinals in a match after a 85 year losing streak. This resulted in a spike in the number of births roughly 9 months later. FC Barcelona's win in the 2009 UEFA Champions League semi-finals caused a 16.1% jump in the birth rate in Barcelona 9 months later.<sup>a</sup> This was directly attributed to the victory. On Nov. 2nd 2016, the Chicago Cubs, a Baseball team, won the world series after a 108 year losing streak.<sup>b</sup> There was a lot of celebrations in the city. As a result, roughly 40 weeks later, there was a massive increase in births.

<sup>a</sup>Jesus Montesinos, et al., "Barcelona baby boom: does sporting success affect birth rate?" *BMJ* vol. 347, 2013, <https://doi.org/10.1136/bmj.f7387>,

<sup>b</sup>Marwa Eltagouri, "A Cubs World Series baby boom? Some parents and hospitals think so," *Chicago Tribune*, 2017, [www.chicagotribune.com/news/breaking/ct-cubs-world-series-baby-boom-met-20170726-story.html](http://www.chicagotribune.com/news/breaking/ct-cubs-world-series-baby-boom-met-20170726-story.html).

When we are talking about intergenerational ethics, we do run into a bit of a mess. Though Consequentialism does have the better methods for handling these sort of cases, it falls flat in a few regards. The reason is that, for all of the ethical theories we have discussed, there is always an individual, or group of persons, which is/are being affected. Keeping with Consequentialism, since it has the easiest time handling these cases, we will say that an action is wrong only if it makes the well-being of a person/group of persons worse than otherwise. For example, when we are talking about current people, we say that something is wrong when they are in a worse situation than some other action which could have been taken. When thinking about future generations, sometimes, the actions we take directly cause them to exist. Those actions, so long as their lives are better than not existing, can't be said to make their lives worse off than otherwise. But, we still, sometimes, want

to say that the acting generation did something wrong. This is a paradox, each line seems intuitive, but in combination, they can't work. Put clearly, we have The Non-Identity Problem.

## The Non-Identity Problem

As I said in the previous page, there are three, general, powers which the current generation has over the future generations which they don't have towards us. These powers, collectively, lead to a uniquely difficult situation to think about morally. But, the third power, the power to bring them into being, generates a very complex and mind-bending problem, especially for certain kinds of consequentialists. The Non-Identity Problem

An action is wrong only if it makes a person (morally relevant being) worse off than otherwise.

An action which brings a person into existence such that they couldn't have been made otherwise and their life is at least marginally better than never being born, can't have made them worse off than otherwise.

There are some actions which bring such a being into existence and are still wrong.

Since this is a paradox, each of these lines needs to seem intuitive. So, I will give some examples or arguments for the first two and in doing so, show that the third line is intuitive, just makes sense or seems true.

### **An action is wrong only if it makes a person (morally relevant being) worse off than otherwise.**

This is the first line of the paradox and to make it seem intuitive, I could reuse tons of the examples which I have already given in the cases concerning Utilitarianism (consequentialism), but for this one, I have chosen to give another example. This example is quasi-historical, as in all of the events took place, but

there were other factors involved (which, frankly, only make the actions more wrong). This particular line, however, does have an element of a person-oriented nature to Ethics. It's basically saying that there needs to be a person (current or future) who is made worse by the action for it to be wrong. If there isn't a person (future or otherwise) made worse off, then the action isn't wrong.

### Historical Example (Ethics concerning contemporaneous peoples)

George Forde<sup>a</sup> was a farmer in Ireland in 1845. Due to the oppressive practices of the British, he and his family survived on a regular potato crop and the milk of a single cow. This was quite common for his region and in fact, this diet made the people in his region healthier in several regards than their British counterparts. Then the potato famine struck. This was a blight which rotted the potatoes and made them inedible. Removing almost all of the food sources in the region. Other countries and regions throughout Europe were also affected, however, due to relief measures, it did not escalate very far. The British, for a time, did similar, and the Irish were doing quite well. But, a few years into the blight, the British powers saw this as an act of God and removed all support and relief measures.<sup>b</sup> This led to Forde contracting diseases due to malnourishment and starving to death, like approximately 1 million others.

---

<sup>a</sup>Yes, George Forde was a real person, in real life, he was a preacher.

<sup>b</sup>For a more rounded explanation for the ending of support, the British Government saw this, in private, as a way to reform the moral character of the Irish people and used a sanctimonious cover to end the aid (which was, through convoluted machinations, paying for itself).

The British ending their relief programs and aid lead to mass starvation and death. They could have, easily, continued those aid programs and saved more than a million people from a slow death. The vast majority of people will say that the British were wrong in ending their aid. This is, in part, because we know the results of it looking back and, even in that time period, they could have looked at how the other countries were handling the blight. We can generalize this, as we have done in other parts of this class, to any action, which, if it follows, makes this line of the paradox seem correct.

**An action which brings a person into existence such that they couldn't have been made otherwise and their life is at least marginally better than never being born, can't have made them worse off than otherwise.**

I know that this line of the paradox is a bit long and seemingly convoluted, but this is because I need to give some examples for it to make sense. Essentially it is saying that a person can't be made worse than otherwise by an action which was necessary for their creation, so long as their life is better than never being born. Because of the precariousness of a person's existence, as can be seen in my above examples, there are several events (and the actions/situations which caused them) which need to go just right for any person to exist. However, if those events don't happen just right then they will not exist. This means that if we choose to take a different route than the one which will result in a person, their 'good life points' would be 0, rather than whatever they would have been had we chosen to take the other route which would have made them. So long as their life is better than 0, they could not have been made worse by the choice. Take this as an example:

## Bio-dome Bob

Bob is a person living on Earth in a bio-dome in the far off future, breathing recycled air, unable to go outside because the O-zone layer has long since been whipped out. He eats rationed food because farming is impossible in this environment. Earth has essentially become Venus. Bob's life is not the best, but there are enough pleasures to make it better than non-existing. In this future, the past generations chose to continue with the non-green policies, not caring about global warming or the environmental effects. Bob's ancestors, in particular, were coal-mine executives, who were able to ensure that their kids meet certain people because of the profits. Bob could not have existed if the green policies were enacted.

We can give Bob's life a score, this is an arbitrary choice which we can use as a metric for later, Bob's life has 10 good-life points. In order to exist, the previous generations needed to not enact those green policies. If the previous generations had gone with the green policies, then Bob's good-life points would have been 0 (because he would not have existed). So, since  $10 \geq 0$ , Bob's life is not worse than it would have been otherwise. As more generations take place, I can generalize this to all people in the future because of Chaos Theory and things like the Butterfly Effect. This means that for an action which makes a person who couldn't have been otherwise and has a life better than non-existence, can't have made that person worse off than otherwise. There are some actions which bring such a being into existence and are still wrong.

This is our third line to the paradox and it seems like the most intuitive of the bunch, but maybe less intuitive than the first line. For this one, we can look at several examples to justify it. But, really, it's just the underlying intuition behind things like wanting to prevent global warming and why we would think that the previous generations in the Bio-dome



Bob case did something wrong. To start on potential ways out of this paradox, take this case which is based on my Bio-dome Bob case:

### Sunshine Sally

Sally is a person living on Earth in the forest 100 years from now, breathing air from the trees, happily playing outside. She eats food grown from her garden and can drink the water from the tap. Sally's life is awesome. In this future, the past generations chose to change their policies to green ones, caring deeply about the environmental effects and global warming. Sally's ancestors were pioneering wind-farmers, the move to green policies caused them to make a bunch of money and be able to send their kids to different places and meet different people.

We can say that Sally's life is, roughly, 10x better than Bob's. If you were to choose which life you would like to be born into, you would choose one like Sally's over one like Bob's. In this case, we will give Sally's life a score of 100 good-life points. But, same as Bob, if the previous generation had chosen differently, she would not have existed. And in this case, she would have had a good-life score of 0, because she was never born.

### Two Potential Ways Out (there are problems with each)

If we take Sunshine Sally and Bio-dome Bob as our token cases of people in these two far distant futures, we can't say that the people individually were worse off than otherwise, we can only say that they are better off than otherwise. But, which do we choose? Neither is wrong (given our situation). The trick is

to say that what makes something wrong is not, necessarily, making an individual worse, but rather making the collection, the group, worse. Sometimes this is a group of only one person, in which case it's the same individualistic intuition, other times it's more than one. So, for this, we are going to reject the first line of the paradox, but not become non-consequentialists, rather say that we need to take the collective well-being.

## Averagism

One way is to take the overall average of the future people and compare the averages. In Sunshine Sally's world, the average person has a better life than the average person in Bio-dome Bob's world, so we can say that this abstract, average person, is worse off than otherwise and that makes the choice the wrong one. But, there is an issue with this sort of account for the outcomes, although it works for cases like Bob's and Sally's, it gets the wrong answer for more down-to-Earth cases, like those concerning inequalities.

Suppose that the average across the board for life-points is 100 (Sally's life) in some future, but this is because the top 1 percent of the population has all the points, while the lower groups have very few, if any, points, maxing out at 10 (Bob's life). But it's a far worse world than one where everyone in fact had 100 (Sally's world). To drive this home, imagine that you needed to choose which world you would want to be born into, by chance. Would you take the bet on a world with a 1% chance of having a good life or a pretty close to 100% chance of having a good life?<sup>a</sup>

---

<sup>a</sup>This case/thought experiment is a variation on John Rawls' Veil of Ignorance thought experiment. I generalized his thought experiment to concern future generations.

The vast majority of people, if they are sensible, would not

choose to take the 1% bet, that's just crazy. So, we can say that Sally's world is better. But, at the same point, what if I just increased the majority's max by just a few points? This would increase the average by a tiny amount, but make it better than Sally's world. This is an issue because a sensible person still would not take that bet.

## Totalism

Another way we could go about this is to have it just be the net total of the people in each world and say that the abstract group "Future People" is better or worse off because of that. In Sunshine Sally's world, the total is higher than the total in Bio-dome Bob's world, so, that's what makes it wrong. This is called Totalism. But this one too does have its issues as well. For this case, the more people in the world, the better the world would look from an outsider's perspective, even though the individual lives in it are far worse.

Imaging that you have two possible futures, one where A) there are 1 million people each with 100 good-life points (a million Sallys) and another where B) there are 10 million people each with 10 good-life points (10 million Bobs). By totalism, we could not tell the difference between these two futures, and, in fact, if I changed B so that everyone had 20 good life points, totalism would claim that it's the better world, even though everyone, individually, is worse off than in A.

As before, we would not want to take this bet, if I were to give you a choice to gamble on where to be born and I told you to choose which of these two worlds you would 'roll the dice' in, you would certainly choose A, even if I made B have everyone with 20 points rather than 10. This is the sort of issue which we saw with the Utility Monster Objection in the past.

## MODULE XI

*What is justice? What  
is equality?*

# *Part 25: Political*

## *Philosophy*

For this module, we are moving into an area of philosophy known as Political Philosophy. This area of philosophy deals with topics like justice, moral/immoral laws<sup>170</sup>, liberty<sup>171</sup>, punishment<sup>172</sup>, and so on. For my content here, I am drawing much from John Rawls' *A Theory of Justice* (which best frames my personal beliefs regarding these things) and Robert Nozick's *Anarchy, State, and Utopia* (which is almost the exact opposite stance). Both of those works really frame the contemporary discussions in American Political Philosophy, with Nozick's ideas appearing often in conservative side of contemporary political discourse. To put a slogan to Rawls' notion of justice, it would be:

---

<sup>170</sup>It is possible, unless you are a moral relativist, to have laws which are immoral. For example, slavery was legal, but it was not moral. This area of Political Philosophy asks questions like "how would you enact moral rules into legal ones?"

<sup>171</sup>Questions here concern questions about the nature of liberty, freedom, how far does it extend, when should it be limited. This should not be confused with the free-will debate, as that's a question about whether our actions are deterministic and moral responsibility.

<sup>172</sup>Essentially "why do we punish people?", "how much should we punish?", and other such questions.

### Justice is equality

But, if that's the best way to encompass Rawls' views into three words (it is), then we will need to do a deep dive into the nature of equality in order to understand what it is and thereby understand what justice is. In doing this, also, we will see Nozick's view on justice come up from time to time, as it is in stark contrast to this notion.

## So what is equality exactly?

Equality is often framed in terms of sameness. You and I are the same, so we are equal. We have the same job, so we should be paid equally. We are the same in social standing, so we are equal. Equality is often seen as a political goal, something worth shooting for. For example, if we notice a pay discrepancy, then we try to enact things which change that. In this case, you get the slogan "equal pay of equal work." Similarly, people want legal equality, and this is where we see Rawls' notion in its most obvious setting. If two people commit the same crime, we often think that it's unjust for one to get a lesser sentence than the other. People in protests about these things cry out that all people should be equal. Those who argue for equality, of any form, are called egalitarians.

Various intuitions about the nature of morality or other people have lead to the same, or subtly different, beliefs about the morality of equality. In the case of Christian thought, it stems from the idea that all people are equal in the eyes of God. As a result, to have the kind of society which God would want for us, we should all be equal in it. Kantians, though often Christians, believe that, due to rationality, all people are due equal respect and from that should be equal within a society. Consequentialists, more particularly Utilitarians, tend to think that equality is the best way to ensure everyone's well-being and get that all people should be equal to get the best society possible for people.

In general, we all have a pretty good intuition about what equality is, and this notion is applied in several different ways. For example, you can have numerical equality ( $2 = 2$ ), equality in quality (being equally good), and many others. But, it should be pretty clear that all people are not equal in all things. Some people are smarter than others, prettier, stronger, taller, and so on. Pretty much any comparison you could make between two people will lead to some being more of that than others (in the case of natural, in born traits). It would be absurd for the egalitarians to argue that people should be equal in all things. Such a world would be one of clones, no one in any way better than any others. Egalitarians, therefore, can't be making an argument with a conclusion like that (Rawls certainly does not).

But, despite this, many opponents of equality frame it in that way. They make the claim that their opponents are arguing something absolutely absurd in order to convince the people to vote/decide against them. This is called straw-manning, where you set up your opponent wrongly as to make them easily defeatable. Because of this easy confusion and the absurdity of the general, we must think of equality as being limited to various contexts. So, "all people should be equal" is not some statement about a world of indistinguishable people, but rather is a statement about equality in a particular context. With that being said, there are three places where people have argued for equality in sensible ways:

1. Equality of the distribution of money
2. Equality of employment opportunities
3. Equality of political power

In order to make these sorts of equality sensible, various people have limited them even further, and in the case of justice as equality, have altered to scope of the equality, or given special exit-clauses.

## Equality of Money

In *Star Trek* as well as *The Orville*, the future is portrayed without money. There are no paychecks for work done, there's no paying for food and water, and there's no paying for luxuries. The reason, explained in the shows, is that technology has removed the necessity of having money gauge rank, merit, or worth. Something, or several things, has replaced money (even digital coinage, like those cryptocurrencies) as the driving force to get things done. Basically, if we were to reach the point where we had something like the matter replicator from *Star Trek*, then either money would have already disappeared or it soon will.

But, as it sits now, we don't have that sort of technology, we need some kind intermediary to exchange goods.<sup>173</sup> Equality of the Distribution of Money can be seen as a rather extreme view, way too utopian, but that has not stopped people from arguing for it and various lesser versions of it are found around the globe, as we will see. This is the stance that all money in a society should be equally distributed among all of the adults in the society. Everyone gets the same income, so to speak. Some have put this as a universal income regardless of the working status. But, as always, we need to look at the reasoning behind the idea to see whether it holds water. There are, in general, two lines of thought to get to this idea:

## Utilitarian Considerations

If you like a Utilitarian-style of thinking, then we have that having an equally distributed basic income will maximize the well-being of that society. This is because, as we have seen time and time again, and we will see later (after the discussion), massive income inequality can and does lead to an overall decrease

---

<sup>173</sup>For a great look at the history of money (and why simple barter systems just don't work), check out this series of videos from Extra Credits [link and citation needed]



in well-being. Sure, it's good for those who have a lot, but the state of those with little far outweighs that. People often think that others have the ability to rise up, given enough 'gumption', but this belief is caused by the exact same fallacy which we saw in the last module concerning harming animals and serial killers.

Rawls used this kind of reasoning in *A Theory of Justice*. Though Rawls and the vast majority of Consequentialist Political Philosophers don't go this far, it is within the ballpark of possibility. Rawls and others claim that justice does entail equality of this sort, but with a massive caveat. Income inequality is just only if the population of the society, over all, is better off with that inequality than without it. Put into a more practical case, this means that if you are rich and you want a tax-break, then you will need to prove that you having more money than other people makes their lives better, not just yours. Or, in another example, reducing the taxes on the wealthiest of us would actually need to create jobs.<sup>174</sup> Putting this caveat on the equality of money does give you an even stronger Consequentialist position, but it still falls into the issues seen in the next page.

## Kantian-style Thinking

175

If you prefer a more Kantian way of thinking about these cases, then you might like this way of getting to equality of money. But, disclaimer, Kantians might not like this line of thinking. It is perfectly within their ability to deny the morality of distributing money, in that it would be taking a person's

---

<sup>174</sup>I used to use an example involving the Boeing Company for this, how their tax-breaks did not create jobs, but actually killed them, but I got too much recoil from it.

<sup>175</sup>This entire section on Kantian style thinking is up for revision, as an interesting paper has come to my attention concerning Kant on the redistribution of money, which is radically different than how Nozick portrays it.

property without their consent. Kant, himself, thought that so long as the tax was applied fairly, as in a flat tax rate, then it would be fine. But the kind of taxation necessary to ensure equality of this kind would be a progressive tax rate, which would not be fair, per Kant. All people, morally, are equal and should be treated as such. From this moral equality you can get to a social equality. Meaning that the moral society is one where everyone is socially equal. What we see in the world is that money is expression of status in the world, so to make people socially equal, they would need to be monetarily equal.

## Equality of Money (Problems)

### Problem 1: Impractical and Short-lived

It should be clear that equally distributing money among all adults in a society is a logistical nightmare. There are many practical problems involved. Many people have claimed that whatever justice entails must be within practical limits, it can't be so utopian. So, justice must not entail this kind of equality because it's just so impractical. But, in reply, morality does seem to, sometimes, call on us to do the seeming impossible. Since justice is an aspect of morality, it's perfectly possible that justice requires such extreme measures.

At the same point, some have pointed to having, rather than a universal equal income, having a Universal Basic Income. In these cases, which are found in several places around the world, you can make more than a certain amount, but you can't make less. Ideally, this amount would pay for a roof over your head and food in your belly. But if you want more than that, you need to work for it.

**Problem 2: Different people deserve different amounts**

Another problem for this view is that, it is claimed, that different people deserve different amounts of money for the jobs that they do. For example, people often think that people doing really gross but necessary jobs, like garbage-people, make more money than other groups because few want to do it and it is necessary (this is not 100% true, but it is a common conception).

The fundamental difference between egalitarians and those who believe in these inequalities is that the egalitarian holds that only small differences are acceptable and those correspond with need. From each according to their ability, to each according to their need. But the other side says that the inequality of any kind is acceptable, not just in need. To each according to their worth.

**Problem 3: Different people have different needs.**

Some people need more money to live than others. A person who needs expensive medical care daily to survive would have a very hard time living in a world with equality in this way. A method for distributing the money would need to include some kind of respect for basic individual humanity. Some kind of, in this case, universal health care, which is argued against as a step towards communism.

**Problem 4: It is not right to redistribute.**

The best arguments for this can be found, not in Ayn Rand, but in Robert Nozick. This is that all of society should be structured in such a way as it ensures you keep what you make. "Taxation is theft" is a more contemporary way of saying the core point of Nozick's work "taxation is slavery."

## Equality of Opportunity

This is NOT the stance that all people should have the same chance to get a job that they want, though it is sometimes given in that way. To see why saying all people should have the same shot at any job, take the following examples:

Suppose that I was applying to be the president of this college, the highest, top-tier position. I do not have any experience in that kind of work, aside from working with some committees here and there nor did I take any classes on college administration in school. Also suppose that at the same time another person was applying for the job (it can only go to one of us). This person took classes in college administration, they have experience in the field, and worked closely with the previous president, so they know the ins and outs of this particular college. If equality of opportunity meant that all people had the same chance of getting any job if they apply, then this would mean that the hiring committee would need to flip a coin to decide who got the job.

Similarly, suppose that two people had recently graduated from medical school, specializing in brain-surgery. One of them spent their summers working as a fugu-chef, getting some renown in that regard, the other spent their summers working as a mechanic on diesel trucks, also getting some renown. Assuming all else is equal, who would you rather have performing delicate surgery on you?

But flipping a coin in this way is not how things should be done, and it should be clear why through looking at the second example. In the second case, it would be really strange for you to pick the mechanic. The delicate and precise hand-eye coordination required to prepare and serve blowfish (which is highly poisonous, if the smallest error is made in the process) is a relevant factor in making the choice.

So, equality in opportunity can't be too extreme. With this in mind, we will say the following:

Equality of opportunity is the stance that two people should have the same shot at a job given their abilities for the job.

No factors unrelated to the job should be included. Sometimes factors unrelated to their education or direct experience can be included (in the case of the mechanic vs fugu-chef), but those are related to the job. So, in the case of being president of the college, I should not be given the same chance as someone with experience, because I am unfit for the position. That being said, if two people are equally qualified for a position, then the just thing would be to flip a coin.

Failure to adhere to this sort of impartiality is most often seen in cases of nepotism. Nepotism, literally, is the favoring or giving preferential treatment to friends and relatives when making a hiring decision. Often, in such cases, the treatment is so preferential that people wholly unqualified for a position are given it solely because of their relationship with the person making the choice. Doing such a thing is a very dangerous decision, because it's likely that they will do worse in the position than someone else. I believe it was in *The Princess Diaries 2* where it was said "Nepotism belongs in the arts, not in plumbing".

On the other-side, a common way to promote equality in opportunity given ability is found in the real world through Reverse Discrimination practices. Reverse discrimination means actively recruiting people from previously underprivileged groups. In other words, reverse discriminators deliberately treat job applicants unequally in that they are biased towards people from groups against which discrimination has usually been directed. The point of treating people unequally in this way is that it is intended to speed up the process of society becoming more equal, not only by getting rid of existing imbalances within certain professions, but also by providing role

models for young people from the traditionally less privileged groups to imitate and look up to. Take the case of philosophy professors for example:

In 2003, 16.6% of the full-time philosophy professor positions went to women, but women made up 27.1% of the applicants, and there were no women of color at all. In order to encourage more women to pursue this field, especially since there is a history of them doing well, women applicants are starting to be treated better than their masculine counter-parts. This has lead to multiple departments around the country being mostly women, for example the department at Pierce College in Washington State (as of 2018).

You will notice that the vast majority of the work read in this class was by men. This is in part because there aren't many readings appropriate for this level on the relevant subjects written by women. But, adding in more woman philosophers would be a change of pace for the standard classes (I happen to know that even PHIL&101 classes taught by women have mostly masculine authors). In selecting my readings for this course, if I had given preferential treatment to works written by members of underrepresented groups, then I would have been engaged in reverse discrimination.

## Reverse Discrimination (Problems)

Though reverse discrimination is seen today in the workplace as well as elsewhere (such as, this kind of thinking is applied in cases of political decisions, like who to vote for), there are some issues worth discussion. Most of the time, the issues which arise from applying this kind of procedure boil down to one, or both, of these two, but there are other issues which can, and do, appear from time to time.

## Problem 1: It is actually anti-egalitarian!

Remember what I said previously about nepotism. This is giving preferential treatment to friends and family, when making a selection. The applicant's relationship with the reviewer/hiring person is irrelevant to the job, so nepotism just can't fit into equality of opportunity. At the same point, reverse discrimination just can't fit into equality of opportunity. Put into a slogan, this issue can be summed up as:

Reverse Discrimination is still discrimination

The aims of reverse discrimination may be egalitarian, but some people feel that the way it achieves them is unfair. Remember what I said about this stance, equality of opportunity does not allow for any features not related to the job to be taken into account in the hiring process. For example, in the real world, I am not allowed to have my picture in my CV<sup>176</sup> because what I look like is a irrelevant factor. Similarly, if I had given favoritism to women philosophers in choosing my content, I would have been bringing in a factor which is not relevant to the quality of the course which I want to provide and create. For a staunch egalitarian, a principle of equality of opportunity in employment means that any form of discrimination on non-relevant grounds must be avoided. The only grounds for treating applicants differently is that they have relevantly different attributes. Yet the whole justification of reverse discrimination rests on the assumption that in most jobs such things as the sex, sexual orientation, gender, or racial origin of the applicant are not relevant.<sup>177</sup>

---

<sup>176</sup>A CV is a kind of resume used in academic and more brainy lines of work. It tends to be far more in depth, but far more dry.

<sup>177</sup>Such identifying characteristics fall into what Rawls calls "non-moral factors" when discussing his test for how just a society is. In his case, reverse discrimination is bringing in "irrelevant factors".

## **Problem 2: Resentment**

This issue is one which is found quite often, though I am happy to say that I have never felt this way. Reverse discrimination is supposed to make access to various professions more evenly distributed across the population, but sometimes, this is not the case, in fact, it can have the opposite affect. Sometimes, in order to reach certain standards of equality, the groups will have no choice but to hire a person who is simply not qualified for the position, or intentionally not hire the best person for the job because they are not from a disadvantaged group. Those who fail to get a particular job because they happen not to come from a disadvantaged group may feel resentment against those who get jobs largely because of their sex or racial origin. Similarly, if a person gets a job because of their status and they fail at it, this can lead to other members of that group thinking that they, too, will fail.

## **Problem 3: It attacks the symptom, but not the cause.**

Another issue with this is in how important it is. When we look at why certain groups are underrepresented in various fields, we see that it's actually the symptom of further, more deep seeded, issues. These issues include residues from racist economic and housing policies, backwards education spending measures, rising cost of an education, and many more. Some of these issues are more historical, but the results of them still ripple into today. Reverse discrimination practices do not treat these issues, rather they only treat the symptoms.

In order to truly resolve these issues in a fair and just way, and thereby get a more representative distribution of people across fields, we would need to radically change our mindsets about economic and educations policies. For example, we would need to make public colleges and universities tuition free (at the very least), better standardize and evenly distribute the qual-



ity of education across the country, and encourage and implement public works projects in the poorer communities to provide quality jobs in those regions. These things are all doable and do not involve anti-egalitarian practices.

## Equality of Political Power

Another area where equality is argued for is in the sphere of government. It is said that all people should have the same impact in political decisions or have the same ‘volume’ when it comes to their voices being heard. In this case too, there is the potential for straw-manning, as we have seen before. The opponents to this kind of equality will claim that it means that a new born has the same political voice as a well-informed adult. This, frankly, is just wrong, and it should not be hard to think of why. Often, in cases like this, the right to make the choice falls to the guardian, meaning that if you have a ton of infant-children under your care, you get to vote on behalf of each of them. This is not what is meant. Rather we need to limit it:

Equality of political power is the stance that all people, who are capable to make informed decisions for themselves, should have an equal impact in the decisions made by their government.

This is called a democracy. Often, democratic processes are done through voting, and in these cases, all people (having the relevant abilities) should be entitled to vote. This is different from a republic where the people have an equal say in electing people to make decisions for them. Normally, the term democracy has two different senses. The first emphasizes the need for members of the population to have an opportunity to participate in the government of the state, usually through voting. The second emphasizes the need for a democratic state to reflect the true interests of the people, even though the people

may themselves be ignorant of what their true interests may be. This is freedom of speech, as seen in the voting process.

## Direct Democracies

Some of the earliest forms of democracy were direct democracies, these are where those eligible to vote have a say in each individual issue. This is like a Reddit-kind of government. For every issue or decision, all people can, and should, vote on it. The solution which gets the most votes is the winner and that is the course taken. Typically, these are only feasible with smaller societies or when there are not many choices to make. For larger societies or for those with many complex issues, there are practical limitations. For an interesting, humorous, Sci-Fi look at this, see the episode Majority Rule of *The Orville*. There are other issues with this, such as the people are not experts, they can and often do make the wrong choice.

## Representative Democracies

In a representative democracy elections are held in which voters select their favored representatives. These representatives then take part in the day-to-day decision-making process, which may itself be organized on some sort of democratic principles. This streamlines the process a bit and helps remove some of the worries which appear in direct democracies. It also removes a lot of the pressure on you to make the decisions which affect others, the person elected gets that pressure. Essentially, in a representative democracy, the people vote and elect individuals from their midsts to make the day-to-day government decisions on their behalf.

There are several different ways in which such elections are conducted. Some demand a majority decision; this is where the representative must have gotten at least 51% of the votes. For example, in these sorts of systems, we can have cases like these:

Michael, Annie, and Billie are all running for the same senate seat. Michael gets 10% of the vote, Annie gets 30%, and finally Billie gets 60%. In this case, Billie wins the election and takes the office. But, running for president, we have John, Paul, George, and Ringo running. Paul gets 20%, John gets 10%, George gets 21%, and Ringo, because of his acting in a hilarious slap-stick comedy, gets 49% of the vote. Because none of them got above 50% of the vote, none of them take the seat.

In such cases, there needs to be additional rules for what happens when none of the people running get the necessary majority.

Others, such as the one used in Britain, demand that the representative merely got the most votes. This removes the worry about needing to get some absolute majority and the need to have a procedure for when one is not reached. This leads to cases like these:

Murdoc, Tudee, Noodal, Hobbs, and Phoo are all running for a seat in Parliament. Phoo gets 10%, Noodal gets 5%, Murdoc gets 20%, Hobbs gets 25%, and Tudee gets 40%. Since it's merely based on who got the most votes, Tudee will take the seat in Parliament.

Representative democracies achieve government by the people in some ways but not in others. They achieve government by the people in so far as those elected have been chosen by the people. Once elected, however, the representatives are not usually bound on particular issues by the wishes of the people. Having frequent elections is a safeguard against abuse of office: those representatives who do not respect the wishes of the electorate are unlikely to be re-elected. Most democracies today are representative democracies.

## Democracies (Problems)

Yes, we live in a representative democracy here in the United States. Luckily, we live in a society where we can and do question the nature of our government. Plato, for one, was opposed to democracies, not just because that sort of system killed Socrates, but also because he thought that the people can't make good decisions for themselves about what's good for them. Similarly, several thinkers throughout the world have found various issues with democracies. Here we will be covering three of them.

### Problem 1: It's an illusion

Some theorists have attacked these forms of democracy as providing a merely illusory sense of participation in political decision-making. They claim that voting procedures won't guarantee rule by the people. Some voters may not understand where their best interests lie, or may be duped by skillful speech-makers. And besides, the range of candidates/choices offered in most elections doesn't offer voters a genuine choice.

In the US (and this is true), various studies have been conducted concerning the political views of the average American. These are done by asking questions about various policies or giving them thought-experiments (the latter being the more accurate). They find that the average American tends to be center-left when it comes to policies. But if we apply the same metrics to the candidates being offered to the people, we find that the farthest left they go is center-right. This limited choice has slowly moved the policies and choices made to the right, when they should, according to the people, be at least, left of center.

It is hard to see why this sort of democracy is so praised

when it typically amounts to choosing between two or three candidates with virtually indistinguishable political policies. Similarly, when it comes to policy decisions, often the choices put forth are merely a yes-no choice on a single law which the people, elected or otherwise, would need to fully understand in order to make a decision on, and they, simply, are not given enough time.

## **Problem 2: Voters are not experts**

Plato and others have pointed out that sound political decision-making requires a great deal of expertise, expertise which many voters do not have. Thus direct democracy would very likely result in a very poor political system, since the state would be in the hands of people who don't know what they are doing.

The captain, not the passengers, should steer the ship.

A similar argument can be used to attack representative democracy. Many voters can't judge how good a particular candidate is for the job, since they don't know what they are doing in the case of political policy. People tend to choose their representatives on the basis of non-relevant features, such as how good-looking they are, they are the same gender as me, they speak in a way I can understand (5th grade reading level), that they aren't a politician (which should be a disqualifying factor in some cases), and so on. This could be seen as a form of nepotism, or at least something like it. Remember that nepotism involves taking in irrelevant features (in this case, relation to you) into the selection process. Here, they are bringing in irrelevant features as well. Similarly, another irrelevant feature is the political party which the person belongs. A person may run as a Democrat, for example, and be totally inappropriate for the position. As a result, many excellent potential representatives are deemed untouchable by the populous, and many unsuitable ones get chosen on the basis of irrelevant qualities they have, of which I am sure you can think of examples.

However, this evidence could be turned around and used as an argument for educating citizens for participation in democracy, rather than abandoning democracy altogether. And even if this is not possible, it may still be true that representative democracy is, of all available alternatives, the most likely to promote the interests of the people.

The major problem—one of the major problems, for there are several—one of the many major problems with governing people is that of whom you get to do it; or rather of who manages to get people to let them do it to them. To summarize: it is a well-known fact that those people who must want to rule people are, ipso facto, those least suited to do it. To summarize the summary: anyone who is capable of getting themselves made President should on no account be allowed to do the job.

-Douglas Adams

### Problem 3: The Paradox of Democracy

For this, I will start with an example and then I will explain how this leads to a contradiction in my beliefs. I will need to, at the same time, say that we need to do something and that we shouldn't do it.

Suppose that I believe that eating meat is barbaric and should never occur in a civilized state (like Plato). If it came up for a vote, I would vote against meat-consumption. Yet, suppose that it did and I voted against it, and the majority decision is that it should be the case that people are allowed to eat meat.

I could replace eating meat in this example with just about any moral belief which I have. In any such case, I am faced with a paradox. If I am committed to democratic principles, then I believe that the majority decision (or the decisions of

the appropriately elected person) should be enacted. But, as a person, I will have certain strongly held moral beliefs about certain things. If I end up in the minority, the losing side, of a democratic vote about what I believe (for example, that refugees should be allowed in the country, that various income tax exemptions should be closed, etc), then I will both believe that it should not be and that it should be. In the case of eating meat, in the example, I would, at the same time, believe that eating meat should not be allowed and believe that it should be allowed.

# *Appendices*



## *Common Biases*

For this section, we will be covering around 16 of the most common biases which people experience, but there are loads of other ones. For example, one which I have experienced is a rhyming bias. This is where you think something is true or accurate because it rhymes. This is called the Rhyme-as-Reason Effect. This is a compound of a few different facts and biases. It's pretty well been shown that people have a tendency to remember things better when they rhyme (why do you think commercial jingles and slogans work?). This means that if something rhymes, then we are more likely to remember it. Then there's the additional bias, which we will cover, which is that we think something is true because we believe/remember it. For example, "if the glove doesn't fit, you must acquit", or this children's rhyme:

### **Mrs. O'Leary's Cow**

Late one night, when we were all in bed,  
Old Mother Leary left a lantern in the shed,  
And when the cow kicked it over, she winked her eye and  
said,  
"There'll be a HOT time on the old town tonight."  
FIRE, FIRE, FIRE!

Many of us know that this is about the Great Chicago Fire,

and many of us now believe that it was caused by a cow kicking over a lantern, mostly because of how easily this song was remembered. However, this is not accurate. Similarly, phrases which rhyme seem to trigger this bias without needing to be committed to memory. According to a couple studies, people tend to believe that something is more accurate when it rhymes. For example "woes unite foes" vs "tragedy unites enemies", "an apple a day keeps the doctor away" vs "an apple a day keeps you healthy", "beer before liquor, never sicker", and so on. We can avoid this bias by simply looking closer at rhyming phrases and look at why they were used, they may be there for persuasive reasons. Every bias has at least one fallacy associated with it and the fallacies get their power (we think they work or they go unnoticed) because we are under the influence of the bias. Noticing and acting against the bias is a great help in all areas of life.

## Confirmation Bias

We all have some strongly held beliefs and we like to think that these are founded on good reasons, but sometimes we accept reasons based on our previously held beliefs rather than beliefs based on reasons. Confirmation bias is where we look harder for and more easily accept evidence/reasons which confirm our beliefs when investigating a stance. Similarly, we don't look as critically at confirming data as we do disproving. At the same time, often, this bias appears when we don't look as hard for and/or don't accept evidence/reasons which disprove our stance. I am sure you have heard of the flat-earthers or the anti-vaxxers. In both of those cases, the believers readily seek out evidence which supports the idea that the Earth is flat or that vaccination causes Autism (or some other nonsense), but are resistant to the overwhelming amount of disproving evidence. Take Anti-vaxxers as our prime example.

In 1998, Andrew Wakefield and a colleague published a paper which claimed to show that the MMR vaccine caused dangerous levels of certain proteins in the blood, which went into the brain causing Autism. To support this claim, Wakefield described 12 children who had developmental delay who received the MMR vaccine and were, 1 month later, diagnosed with Autism. Groups who already had the belief, for various reasons (most, if not all wrong), that vaccines were dangerous, accepted this finding whole-hog without digging much deeper. But this is flawed on several accounts. If you dig deeper into the study, you realize that children are diagnosed with Autism around the same time as they would receive the MMR vaccine (the evidence appears at around that age). So, this does not show causation. To show causation, there would need to be a large sample size with an equal mix of children who would not receive the vaccine and those who would and then there would need to be a remarkable increase in the cases of Autism in the vaccinated children. This was not done and subsequent studies have shown that this would not happen. Additionally, looking at the 12 children which Wakefield mentioned, developmental delay is one of the early signs of Autism (though, I will say, that developmental delay does not always mean Autism), so that example does not actually support his case.

But, all that being said, the people, with the preconceived notion that vaccines were in some way bad, did not do that digging, they just accepted the finding. Otherwise, I would hope, reasonable people would do the digging and learn more about the study, seek out ways in which the experimental parameters could have been wrong, but this confirmation bias derailed the train. The fallacy associated with this should be easy enough to see, it is where we put more weight on evidence or an argu-

ment because we agree with the conclusion or we think that an argument is less strong than it is because we disagree with the conclusion.

## Self-Interest Bias

This sort of bias is pretty easy to see in other people, especially when their behavior affects others as well as themselves. Self-interest bias is where we act or believe something merely because it benefits us (the individual) in some way, without considering the effects it might have on others. If you ever jump to a conclusion which is beneficial to you without thinking about others, then that bias has likely triggered in you. This bias causes us to overlook the interests of others and, thereby, is not a reliable guide to what is the right thing to do in most situations (involving others). Take this example:

A young man (adult) has lived and worked (on his own) in another state from his parents for a year and, at the start of the second, started going to the local state university to get his degree. His father, when doing his taxes (for the first year of the son being away), listed the young man as a dependent. Doing so saved the father approximately \$2,000. However, when the young man applied for in-state tuition at the university, they claimed that he was a dependent of an out-of-state person, and thereby did not qualify, doubling his tuition from \$10,000 per year to \$20,000 per year.

The father in this case was acting in a self-interested way. If the father had confirmed with the son what listing him as a dependent would do and thought of him, then they either could have come to some sort of arrangement ("hey pa, I'll just pay you the two grand once I have myself established") or the father could have just ignored that sum. The bias made him

gloss over the effects on others.

This is not to say that we should always act altruistically, that would need to be argued for in a different way (if accurate). Sometimes the right thing to do will benefit you and may even be at the expense of others. However, the bias comes into play when your reasoning does not take their interests or the effects which the actions might have on them into account at all.

## Affect Bias

This sort of bias is found in several logical fallacies. This is where you allow your emotional response to a case or situation to override your logical thinking or your actions. For most people, the idea of the Nazis generate a strong emotional response of hatred and the idea of, say, marriage or love generates a strong positive emotional response. You can trigger this bias in another by using words, tactfully, in your argument. You could also do this to yourself, if the situation is right. For example, suppose that a person is on trial for child abuse. That context alone is enough to cause an emotional response. The evidence against the person could be outstandingly weak, but it would still be an uphill battle for them because the emotional response about such a case will have triggered this bias in your fellow jury-people and thereby give them the gut response of guilty, without even entertaining the evidence. Similarly, in extreme situations, some people have needed to resort to cannibalism to survive. Upon hearing about such a case, your gut will likely twist and made you immediately jump to the conclusion that what they did was wrong. However, this emotional response clouds the reasoning being resorting to cannibalism and makes it harder to see that the actions were for the better of the group as a whole.

This bias gets its power over us because people often think that the strength of the conviction about something (an emotional attitude towards it) is the same as the strength of the

reasoning behind it. Take these two cases:

### **The Origami Frog Killer**

You are the head of a detective team which has been tracking a very violent serial killer. The tabloids have dubbed this murderer “The Origami Frog Killer” after the distinctive calling card, an origami frog placed on the mangled remains of what was once the victim’s chest. Eventually, after months of effort, your team finds them. She committed suicide and left a note mocking the police force and your team. You know that if this news came out, that you failed to bring this murderer to justice, the pain and grief felt by the populace would be extreme. There would be great distrust in the government and police force, rioting, potentially copy-cats, etc. However, there’s a fairly weak-minded individual in your holding cell for a petty crime. Your team is very loyal to you and knows that if you go down, they go down with you. Since only your team has access to the precise evidence needed to get the correct criminal, you can easily alter it, convince the person in the holding cell that they committed the crimes, and thereby frame them for it. This will prevent the grief, distrust, copy-cats, etc. which would have happened otherwise.

Do you frame the person in your holding cell?

### **The Red Rose Gang**

You are the head of a detective team which has been trying to get a very violent gang leader behind bars. You and your team know full well that she has committed very heinous acts, including cop-killing, drug-sales, extortion, slaughter of families and children, etc. She will soon be in court for some of the crimes, but not all. The defense, however, is very crafty, and given the evidence which you could present to the court, it's very unlikely that she will see jail-time, much less the life sentence which justly would come with the conviction. This will allow for the continuation of her activities, and, maybe, embolden her to make even more violent acts. Your team, who think like you, will follow your orders about this case to the letter. It is perfectly within your resources to alter and forge evidence to make the case stronger and ensure that this person would never leave prison. Doing this will prevent the continued violence which her group engages in, clean up her drug-ring, and make people less fearful of such groups, making them more likely to report them. Do you forge the evidence?

For the first case, The Origami Frog Killer, your immediate gut reaction is, likely, not to forge the evidence. We have a strong emotional reaction to the idea of framing an innocent person. For the second case, your reaction was likely in favor of forging the evidence. But, what is the difference? In both cases you are forging evidence and the consequences of doing so are similar (restoring faith in the justice system, preventing death and harm, etc.) Without the emotional response, your answer to the final question should be the same in both cases, either forge the evidence or not. In the second case also, this emotional response can counteract the accepted idea that people are innocent until proven guilty. The trick to avoid this bias is to recognize when your emotions have been triggered and try

to suppress them.

## Selective-Attention Bias

This bias is a bit different. This is where you tend to notice evidence which supports your beliefs and not noticing the ones which contradict the belief. To use Anti-vaxxers as an example again, a person might see all the evidence on both sides of the 'debate' but fail to notice the extremely vast amount of evidence which supports vaccination, turning their gaze instead to the scant few shreds of evidence which oppose vaccination. This will cause them to think that the reasoning is on their side in being opposed to vaccination, but in reality it is not. People who are swayed by this bias don't realize that we experience the world and get information through a filter. Often this filter is tuned so that it lets through the good information, confirming or comforting, and blocks the bad, contradicting or disheartening. To overcome this, you need to realize that you have a 'dog in the fight', so to speak, and pay closer attention to the evidence which your filter tries to block.

## Negativity Bias

This is similar to the affect and selective-attention biases, but it does not, necessarily, involve emotions or a filter. For this bias, the tendency is to place greater weight on the negative information about something and giving less weight to the positive information about that thing. This bias can be most seen in the political arena. Political campaigns are often full of what we call 'attack ads' about the opponent. This information (or disinformation) sticks in the head because of this bias. We place a greater weight on the negative information than the positive and use that to justify our voting. Now, sometimes the negative evidence should be weighted more than the positive, but this is not because it is negative, rather because it has more rele-



vance or pressing aspects to it. Sometimes, too, the positive information needs to be weighted more, for the same reasons. The trick to avoid this is to be impartial about the positivity or the negativity of the information given, and weight them according to relevance rather than by the kind. For example, imagine that a news headline read "Buy N Large CEO laid-off 1000 employees". This case would definitely stick in your mind and you would weigh that heavily if you are under this bias. The next headline reads "Buy N Large CEO creates 1000 new jobs in the affected areas". This directly counteracts the lay-off but, if this bias had its way, you would still think negatively of the CEO.

For another example, suppose, as is the case, that news got out that the Prime Minister of Bhutan refuses to make deals with oil and gas companies to allow them to use his country's land and resources. The news would likely paint this choice rather negatively "the PM of Bhutan refuses to allow companies into the country, stifling job growth". The Negativity Bias would have us remember this more strongly than the news (and facts) that Bhutan is a very small country, with a stable economy, and is the only carbon negative country in the world (not neutral, negative). They have no interest in that kind of economic growth because they already have an energy surplus which they sell to other neighboring countries.

There's good reason we remember negatives over positives, historically, negatives were more of a threat to us and things which would save us if we remembered in the future. But, this hold-over from the past leads us more astray now than it used to.

## Resistance Bias

Every last one of us can be wrong from time to time, no matter who they are or what they are doing. The more training you have in a particular field the better you are at preventing

yourself from being mistaken in that field, but it's still possible. Having others around to catch our errors and correct us helps us improve and become better. However, we should not always be so self-debasing that we always change when we are told we're wrong. Sometimes, the opposing side is wrong and they are the ones who need to be corrected. There is a middle-ground which needs to be found in these cases. Resistance bias is when we are overly opposed to being corrected, even taking deep offense to the idea that we may be wrong. This causes us to miss opportunities to improve and become better. The show "The Big Bang Theory" has some great examples of this, mostly by Sheldon Cooper. In those cases, Sheldon refuses to admit that he was mistaken or just plain wrong in some case. Typically, by the end of the episode, the issue is resolved, but Sheldon's refusal to even entertain the idea that he may have been wrong is the sign of this bias. Think about how you react to criticism? Are you overly passive? Or do you just refuse to accept it all together? If it is the latter, then this bias is a strong one for you.

The trick for this bias is to accept that you could be wrong, be humble. Accept criticism, consider it impartially. But, you need to also recognize that you could be correct. So, when you are considering the criticism, don't just show your belly and give up.

## Belief Bias

This is one which I most often encounter in students when I am teaching about Ethics or arguments concerning the existence of God. Belief bias is where we judge the strength of an argument based on whether we agree with the conclusion, not on the actual logical structure or the supporting evidence. Many students come to college with the 'woke' idea that morality is in some way relative to the culture, that it depends on the beliefs of the culture and that there's no facts about morality (this is,

in part, the next module, and we will, likely, see this argument again). So, what I do is spend around a week with students proving that this stance is both unsupported and, if it were true, actually quite horrifying. After giving several examples of cultures with radically different views about morality than our own (one where infanticide is fine and another where the correct funeral rights is to eat the body), I give the following argument:

### **The Cultural Differences Argument**

Cultures have different views about what is moral.

Therefore, there is no objective morality (it's relative to the culture).

If you are under the influence of this bias, then you would think that this is an extremely good argument. However, this argument is very weak. First off, the argument goes from what people believe to what is actually the case, which is a wrong move on many levels. For example, if I keep the argument structure the same and use the same kind of reasoning, some funny things pop out:

### **The Shape of the Earth Argument**

Cultures have different views about the shape of the Earth.

Therefore, there's no objective fact about the shape of the Earth (it's relative to the culture).

This argument, in form, structure, evidence, and everything else, is the same as The Cultural Differences Argument. If you thought the first argument was great but the second was loony-tunes, this bias got you. Here are two more arguments for examples:

**Brain and Computer**

The brain is like a computer.

They both have complex, inter-working parts and it's very clear when one part fails to function properly.

Yet, it would be ridiculous to claim that a computer didn't have a designer. So, claiming that the brain didn't have a designer is equally ridiculous

**Hammer and Gun**

A guns is like a hammer.

They both have metal parts and could be used to kill someone.

Yet, it would be ridiculous to restrict the purchase of hammers

So, restrictions on purchasing guns are equally ridiculous

Both of these arguments are very weak, if we treat them as inductive, and insanely invalid, if we treat them as deductive. However, in reading them, there's a strong likelihood that you thought that one was better than the other or you thought that both were great. Either way, this bias got you. If you thought the arguments were bad based on the conclusion alone, then this bias got you as well. The key thing about this bias is that you need to take your personal feelings out of the evaluation of the argument. My trick is to replace the thing which I agree with in the argument with something else. So, in arguments about the existence of God, I replace it with 'flying golden banana'.

If the argument still has the same punch, then it works, but if it doesn't, then I know that I fell to the bias.

## Availability Bias

Quick question for yourself (don't look it up): How often do people win the lottery? How frequently do celebrities get together and break-up? For the latter, I don't have the answer, but I would say not as often as you would think. For winning the lottery, it's actually very, very, rare. Lotto agencies spend a good amount of time and effort pumping our minds with examples of people winning and how awesome their lives turned out to be. This makes this evidence very available to us. Availability bias is where we are lazy and use examples which are easy or quickly come to mind. For example, the cases of cultural oppression and one culture forcing themselves on another (Cultural Imperialism) are fairly easy to come by. The easiest of them are the ones where the culture was wrong to do so. You have the actions of Spain in Latin America, the US towards Native Americans, the British in Asia (in general), and so on. These 'low hanging fruit' examples might make you think that one culture imposing its values on another is always wrong. If that's all you did, the availability bias got you. But, there are plenty of other cases where Cultural Imperialism was actually the right call. If you are of certain religious leanings, then you might think that missionaries are doing something good, but they are engaged in this behavior. For a less contentious example, any nation which sided with the Allies in WW2 were imposing their cultural beliefs on another. Going in and trying to introduce civil/human rights in a region also counts as Cultural Imperialism. These examples show that if the information is easy to come by, the conclusions drawn from them could still be wrong. The trick with this bias, how to avoid it, is to recognize how easy the information was to get or how immediately the examples jumped to mind. The easier or the faster, the

more you will need to be cautious about the conclusion drawn. Sometimes, and this is true for all of them, the bias might actually get you the right answer, but this is by luck, not by good reasoning.

## Bandwagon Bias

Come on, man, everyone is doing it. You don't want to be left out, do you? If you were right, then don't you think more would think it, too? The last three sentences are variations of things which you may have heard in the past. Peer-pressure and other things like that work because they are acting on the Bandwagon bias. This bias is where we unconsciously tend to adopt stances, beliefs, or behaviors because many other people have them or they are had by a group which we belong. In football (either soccer or American football), we see this when people support the teams which others in their area support. This is not because, necessarily, the team is actually good but rather because other people around us support them. 'Bandwagon fans' are another example, they switch their support according to who has the most fans and who is winning. In the political arena, people often support the side which their community supports out of fear of being an 'outcast'. Alcohol commercials and tobacco advertisements (when we had them) will often show everyone drinking (or smoking) the product and having a great time. This acts on the bias because it makes us want to belong to such a group and we, unconsciously, adopt the beliefs of that group. Similarly, when I was young, riding the bus to school, the driver would blast country music over the audio system. My fellow students seemed to like it (I, frankly, did not, I grew up on heavy metal and classical). If, in this case, the bias had gotten me, I would come under the stance that country music was good music, not because of the instrumental or lyrical qualities, but because everyone else seemed to like it.

We all have the desire to make friends and belong to a group. It is the sort of beings we are. We evolved as communal and group oriented creatures. Adopting beliefs or stances on those grounds alone is simply not good enough. If you ever find yourself attracted to a stance, regardless of popularity (though this bias works best for the most popular), ask yourself about why you are drawn to the stance. Is it because your friends have it? Is it because it would get you into a group or club? Look at the stance closer and judge for yourself whether it actually holds water. For example, suppose that a young person in the Pre-Civil War South was unsure about the morality of slavery. She might think that because her community thinks it's fine and her opposing it would leave her in financial ruin because of the community would, essentially, kick her out, that slavery must be OK. This would not be her thinking for herself, rather it would be the lazy thing to do. Bandwagon bias, like many of the other biases, is lazy. It pushes the mental work necessary to make a reasonable conclusion off of your plate and on to the the less-reliable masses.

## First-Person Bias

This is one of my favorite biases to talk about because there are some great examples of this at work in history as well as in contemporary times. First-person bias is not where you think information is more reliable because you have first person experience of it (that's another bias). This bias is where you judge what is good for others according to what is good for you. For example, some of you may have heard about the assassination of Julius Caesar. As Caesar rose to power and began to institute his reforms, a splinter political party emerged, The Liberatores. At that time, there were typically two parties, the Populares (the popular ones) and the Optimates (the best ones). The Liberatores were composed of those who believed that Caesar, through the changes, was stifling liberty and hurt-

ing the people. There are several different angles one could take on the reasoning for the assassination, none of which really support the Liberatores as being correct in their actions. What really irked them was Caesar's proposed tax plan, which would cut taxes on the farmers and poor and, in turn, increase the taxes on the wealthy (in order to improve roads, infrastructure, and so on). The Liberatores (changing a few things here for my example) were very rich and this tax policy would hurt them. If it hurts us, they thought, it must hurt everyone else. Upon killing Caesar, they left and found that the masses were far from happy, Caesar's policies would have helped them immensely. The Liberatores failed to see this because of this first person bias.

For another, suppose that you, wrongly, really dislike the rye chips in the snack mixes. A coworker offers some to you, so you, thinking that you were doing right by them, pick out the chips for yourself, saving them the 'gross parts'. They reply to this with something along the lines of 'man, you took the best part!' Here, you evaluated the situation about the tastiness of the snack mix according to your own preferences and did not take their preferences into account.

A common result of this bias is sort of the opposite of the negativity bias. Because of this bias, sometimes, we place a higher (unwarranted) weight on what is good for ourselves and place a lower (unwarranted) weight on what is bad for others. With the Liberatores, they placed a higher weight on what was good for them and did not factor in (or didn't factor in as much as they should have) the negatives on other people. This has resulted, and has continued to cause, many social injustices in the world. Different groups and communities face different struggles and what is good for one group (such as maintaining a status quo) could add to (or continue) those struggles. The main take-away from this is that to avoid this bias, you need to meet each other on the level, put yourself in their shoes, recognize that what is good for you might not be good for them.



## Ethnocentricism and Stereotyping

This is most closely related to bandwagon biases. It is one of the root causes for racism and many of the struggles which we face today. This is the irrational belief that your ethnic group, society, or culture is innately or fundamentally superior to other groups. This, along with certain other biases, causes the person under its sway to see their group in the best possible light while seeing other groups in the worst possible light. Thus the bias will cause you to fail to see some relevant facts on both sides. Ethnocentrism can lead people to think that an individual is guilty based on the color of their skin or their religion. This bias has led to a great deal of suffering which is still happening to this day.

Related to this is stereotyping. Now, we can't hold all of the information about a thing in our heads immediately. We just don't have the memory capacity for it. So, we, naturally, simplify things and place it into categories. These categories tend to be very simple and, for most things, tend to be fairly neutral. For example, if I present you with a picture of a car, you will quickly put it in the category 'car' and be able to make various assumptions about the car based on the information in the category. This is easy and very efficient and, for many things, it will not lead you too far astray. However, sometimes nasty aspects will infect your categories. These might be positive or they might be negative. In either case, it removes the neutrality of the category. This is where we get stereotyping. They are oversimplified and, typically, unflattering, generalizations on members of a group. For a first example, there are many people who play World of Warcraft. Picture in your mind what you think a player would be. For many, they are unhealthy, socially awkward, and otherwise unpleasant. Those are stereotypical features of a WoW player. Ethnocentrism will cause you, also, (if you aren't a player) to paint this group in that negative aspect and re-enforce the idea. For another example, suppose that you have in your mind that professors with certain accents

aren't as smart as professors with a different accent. This will lead you to stereotype the professor upon hearing their accent and thereby start employing the confirmation bias to support that claim (that the professor isn't as smart). This will reinforce the idea in your mind that the generalization is correct and deeper ingrain it into your mind. In Ethnocentrism, we get that the nasty pollutants in the categories are both positive things for your own group and negative aspects for others.<sup>178</sup>

Avoiding both of these things is a great struggle in this day and age. We see regularly the effects of Ethnocentrism and stereotyping, all not good. You need to analyse your categories, see what is in them, and recognize when these features are unwarranted. This is especially true for groups of people. Never include an evaluative judgement in such a category, whether it is the category for your group or the group of another.

## False Consensus Bias

This is another bias which is most often found in the political arena but it can also be seen in other aspects. You may have heard people claim that if you leave a certain bubble, then you will see that very few actually hold the same stance as you. For example, I live in a fairly liberal leaning state, not denying that (given voting patterns and surveys), but people I know claim that if I were to really check the figures, then I would find that most people in this state are actually conservative. I only ever hear this from conservatives and this is an example of the bias which I am talking about here. The false consensus bias is where you think that most people believe the same things as you merely because you believe it. You don't confirm your hunch or check the real data. It is equally possible for left-leaning

---

<sup>178</sup>Much of the content here and explanations of stereotyping came from an application of the theories and information in Mental Files by François Recanatani and Davis Smith's (my) Master's Thesis "Here Be Dragons".

people to fall for this trap as well. For example, suppose that a person believes that most people believe that there should be far more gun control. The bias comes out when they believe this without checking the surveys. The consequences of this line of thinking can be found in the way certain groups receive their news. Some person posts a theory or a 'fact' (may or may not be accurate), this gets picked up by another group, citing them for credibility, and then on and on it goes, leading to an echo-chamber, making people believe that more and more people agree with them. Conspiracy theory forums are a great example of this, same with other propaganda sources. The belief that many agree with you can also cause you to fall into the bandwagon bias which we saw before, and further re-enforce the belief, without actually adding any real evidence to it.

To avoid this bias, you need to be careful about what you think others believe. This solution is similar, in sorts, to the first person bias. Check to see whether many people agree with you, trust surveys and statistics, or, better yet, learn how to create impartial surveys and conduct them yourself. You may find that most people don't necessarily agree with you. This can be good. Figure out why others disagree with you, what is their reason. You might find that you were the mistaken one.

## Externalization Bias

We have all made mistakes from time to time, it's part of the human condition and it's how we grow and become better. Learning from our mistakes is a key part of life. However, some people will look elsewhere for who is at fault, they will not look at themselves. This is the externalization bias. For example, suppose that I got into a car accident after running a red light. I could blame myself, say that I was distracted while driving and grow from the experience. Or, and this is where I would have the bias, I could blame the light, the other drivers, and anything but myself. I could say that I was sure the light was

yellow or green and claim that the light must have switched the other side to green early. How would I grow and learn from my mistake?

Externalization bias is where we, for various reasons, refuse to take on the responsibility for our actions. Something external to us is causing the issues, not us. We make up excuses. We see this both on an individual level as well as in a group setting. For example, suppose that you are in a competition to build a rocket engine, the one which produces the most power wins. The other teams have far more experience in a competition setting and this is your team's first year in the competition (though you have the same level of experience in building engines). When you don't win, you could blame the other teams, claim that the judges were in some way partial towards them, or say that there was some kind of sabotage (without evidence). This would be the externalization bias at work. On the other hand, you can learn from your errors, figure out what made their engine better, how you can improve your designs, and see that you, as a team, grow from the experience.

Sometimes, however, the issue is external to us. Sometimes, something is wrong with the system. So, you should not always place blame on yourself for a failure. The trick is to fully analyze what you did, learn from that, and then, if everything was perfect, blame something external.

## Substitution Bias

We have all encountered questions which are hard to answer or understand. Often, without realizing it, we will replace or omit key words or phrases in the questions with others, making the question easier to answer, thinking that we answered the original. But, often, replacing those key words or phrases in the question makes it a new question all together and makes your answer irrelevant to the original question. For example, take this question and answer pair:

Question What do you think of the nuclear triad?

Answer Nuclear weapons are powerful and devastating in their effectiveness and power. There needs to be safe-guards to ensure that no bad people can get their hands on them.

If you know anything about the US nuclear weapons capabilities, then you know that this answer makes little to no sense. It completely dodges the question. The nuclear triad is a three-part (triad) military structure which consists of ways in which the US can launch nuclear weapons around the globe. It can be from the land, sea, or air. Reasonable responses to this question include things about whether the structure limits the chances of an accidental firing, whether the structure actually ensures the safety of the country in question, whether the hardware/software involved in launching the nuclear weapons is up-to-date or need to be improved, whether the command structure is efficient enough (or too efficient), etc.

But all of those answers require that the respondent know about the details of the nuclear triad and has thought deep enough about the structure to actually hold an opinion about it (aside from the hardware/software aspects, I know very little about the structures within the triad, so I would need to do research and form an opinion). Those answers are hard to accurately come up with. Intellectual laziness is at fault again when we use this. By just replacing the word 'triad' with 'weapons' in the question, we get a far easier question which can be very simply answered. But, that question does not relate back to the context or the relevant aspects of the original.

For another example, think about this question answer pair:

Question What do you think about the impact of organic farming on global food production?

Answer I think it's great that people have the ability to buy organic foods and be sure that their foods are safe for their families, without pesticides and the like.

In this case, the question was not about the availability or production of organic foods. Correct responses to this question include a stance about the average cost of producing food, the amount of land required to grow organically (it is far greater), the response of food producers to the wave of demand for organic food, and so on. But again, all of these require a good amount of research and the in depth knowledge about the impacts of organic farming, and it requires you to also have confidence in that research to form an opinion. All of those are missing from the answer. Again this is intellectual laziness at work.

Avoiding this involves checking your answer before you speak/write it. Did you gloss over a word in the question? Did you miss something which should have been there? I have encountered hundreds of papers now where a student wrote at length about something totally different than what was expected. Always check yourself.

## Anchoring Bias

An anchor in this context, is a fixed point, something which you build off of or around for a given topic. Sometimes, this anchor can be good, a reasonable starting place for building an understanding on a topic. But other times, this anchor could be placed in the wrong spot. This anchor is typically the first thing you learn about a topic and it shapes or colors your opinions/attitudes (whether you accept or reject) new information about it. For example, many people first learn about the origin of the Earth and humans from religious institutions, like churches. In the Christian stories, it involves God creating everything in 7 days. For some, this is their anchor about the origin of the world. The anchoring bias is when one refuses to recognize that their foundational understanding of the topic was mistaken or that this, often mistaken, understanding colors all other learning about the topic. Some Creationist thinkers

could be accused of having this bias. They set their anchor about the origins of humans on their first encounter and refuse to budge from that position, clouding their understanding of the rest of the data. This is often coupled with confirmation biases as well as negativity biases to further dig the anchor in. You place higher weight on this first chunk of information than warranted.

Similarly, this can come in the form of an attitude. Think about a person who shapes their opinions of others based on a core anchor. They could have the idea 'if they belong to the Flat Earth Society of America then they aren't smart, otherwise they're good'. What would you think of such a person? The anchoring bias has given them a far too granular sense of the world. In this sense, the anchoring bias gives an overly simplistic and inaccurate scale to gauge people on.

Avoiding cases like this involves understanding that all parts of your understanding could be mistake. Look at the reliability of the first thing you learned? Did it come from an actually reliable source? Has new information come out to contradict that stance? Treat the first thing you learned the same as you would anything else. Have a good amount of skepticism when it comes to this.

# Glossary

## **argument**

A connected series of sentences, divided into **premises** and **conclusion**. 5, 15, 577

## **coherentism**

The stance in Epistemology concerning justification that all beliefs must be justified, there can be no ‘foundational’ beliefs. For this theory, it is necessary that some beliefs indirectly justify themselves. For this theory, beliefs are justified in how they cohere or fit together. The structure is like a spider web or loosely woven fabric. 311

## **conclusion**

The sentence which an argument is intended to support or prove, often indicated with a conclusion indicator.. 5, 15, 573, 577

## **consequentialism**

The general stance that the morality or permissibility of an action is determined by the consequences of that action. This is a family of theories which all share the basic command to promote the good, though they differ on what the good is and how one ought to promote it. 369



**cosmological argument**

An argument which moves from observations of the world around us, more particularly those concerning causation, to the claim that there must have been a first cause. 224

**deductive argument**

An argument such that the intent behind it is to guarantee the conclusion. There is no ‘wiggle room’. In general, deductive arguments move from broad general principles and to more particular instances.. 19

**dialectic**

the art of investigating and determining the truth of some stance or opinion. 6

**epiphenomenalism**

The stance in that certain physical events can cause mental events and vice versa.. 113

**epistemology**

A field of study within Philosophy dedicated to questions concerning knowledge, belief, and justification. 288, 291

**error theory**

A general claim that all other claims within a field of study or context will always be mistaken, try as we might to say something truthful, we cannot because there are no truths in that area or context. 360

**expressivism**

A general claim that all other claims within a field of study or context are not intended to and do not express anything true or false, rather those claims are expressing emotions, commands, or questions. In ethics, saying something is wrong, according to this stance, is the same as saying

something like ‘boo’ and saying that something is right is the same as saying ‘yay’. 362

### **foundationalism**

The stance in Epistemology concerning justification which states that in order to know something the justification necessary must be structured in such a way as they fundamentally rely on basic, foundational beliefs which are self-evident and do not need further justification. The structure is often compared to a pyramid or a building. 307

### **good will**

The one intrinsic good according to Kant’s non-consequentialist theory of morality. All rational autonomous people have this ‘good will’. It comes from our ability to rationally and freely choose to act on our duties. According to Kantianism, anything which thwarts or hinders either another’s free will or rationality is wrong. 386

### **idealism**

The stance in that the world consists of only one general kind of substance, mental.. 115

### **inductive argument**

An argument such that the intent behind it is *merely* to make the conclusion more likely, assuming the truth of the premises. There is some ‘wiggle room’. In general, inductive arguments move from a collection of particular instances and then use those to support the truth of some general principle. 22

### **infinite regress theory**

The stance in Epistemology concerning justification that all beliefs must be justified, there can be no ‘foundational’

beliefs but also beliefs cannot indirectly justify themselves. The structure is like an infinitely long chain of branching beliefs. 309

### **justification**

The reasons one has to believe something, why they think it is true. 305

### **kantianism**

The philosophical theories proposed and defended by Immanuel Kant. 386

### **knowledge**

Justified True Belief; You have good reason to believe it, you do believe it, and your belief is, in fact, accurate. 304

### **man-made evil**

The pain and suffering and otherwise ‘bad’ things which human persons do through free will. 236

### **maxim**

The reason and goal for which you are acting. This is typically of the form “whenever I am A then I will E”. 394

### **meta-**

a prefix which changes the level of abstraction for the questions in the field, namely by moving it up. This means, roughly, ‘asking questions about the questions in...’. 332

### **monism**

The stance that the world is comprised of only one kind of substance.. 115

**moral relativism**

The stance that there are facts about morality, that is there are standards of right and wrong, but those facts are variable, they change according to the beliefs of culture or individual in question (which depends on the form of Moral Relativism claimed to be true) plural. 334

**natural evil**

The pain and suffering and otherwise ‘bad’ things which are caused by natural processes of the world, not caused by human free action. 237

**nihilism**

The stance that there are no facts in a given topic. In other words, the stance that the facts aren’t ‘real’, there is nothing true or accurate to be said about it. The name for this stance comes from the Latin word ‘nihil’ meaning ‘nothing’. 333

**non-consequentialism**

The general stance that the morality or permissibility of an action is not determined by the consequences of that action. This is a family of theories which all share the basic idea that some actions are just wrong regardless of the good they might (will) produce; typically, these theories are duty-based, meaning that an action is moral if and only if you have a duty to do it and immoral if and only if you have a duty to refrain from doing it. 369

**ontological argument**

An argument which moves from the properties which an object/thing is purported to have to the claim that the object/thing must exist. 230

**ontology**

An area of Philosophy which concerns the study of what exists. 229

**physicalism**

The stance in that the world consists of only one general kind of substance, physical.. 109, 116

**premise**

A sentence in an argument other than the conclusion, often indicated with a premise indicator. 15, 573

**Problem of Evil**

An argument against the existence of God which points to the pain and suffering on the Earth and asks how an all-knowing, all-powerful, and all-good god could allow for such a thing. 236

**realism**

The stance that there are facts in a given topic. In other words, the stance that the facts are 'real'. 333

**relativism**

The stance that there are facts in a given topic but those facts are variable, they change according to the culture, context, or individual. In philosophy, generally, the facts change according to either what the individual believes or the culture in question believes, depending on the form of relativism held. 334

**skepticism**

The epistemological stance that knowledge is impossible either globally (it is impossible to know anything at all) or locally (knowledge is impossible within certain topics). 288

**sound**

A property of arguments that holds if the argument is valid and has all true premises. 20

**substance**

The thing itself, which has properties and can survive changes in those properties.. 110

**substance dualism**

The stance in that the world consists of two general kinds of substances, in the case of the Mind-Body Problem, these are mental and physical.. 109, 110

**teleological argument**

An argument which moves from a seeming purpose, end, or design exhibited by the universe to the claim that there must be a designer of the universe. 215

**utilitarianism**

A form of Consequentialism which says that the morality of our actions is determined by the amount of happiness caused and suffering prevented. This is a form of Direct Consequentialism which takes happiness as the intrinsic good and suffering as the intrinsic bad. 372

**valid**

A property of arguments where the truth of the premises guarantee the truth of the conclusion; i.e. it is impossible for the premises to be true and the conclusion false. 19

[nonumberlist]

# Bibliography

- 3Blue1Brown. But what is a neural network? Chapter 1, Deep learning. *YouTube*, Oct. 2017. [www.youtube.com/watch?v=aircAruvNjK](https://www.youtube.com/watch?v=aircAruvNjK).
- Appiah, Kwame Anthony. "What will future generations condemn us for?" 26 Sept. 2010, [www.washingtonpost.com/wp-dyn/content/article/2010/09/24/AR2010092404113.html?hpid=opinionsbox1](https://www.washingtonpost.com/wp-dyn/content/article/2010/09/24/AR2010092404113.html?hpid=opinionsbox1).
- Argument Clinic*. 1972. Writer John Cleese, episode 29, BBC1.
- Augustine, Saint. *The Confessions*. Oxford UP UK, 1990.
- Austin, J.L. "A Plea for Excuses: The Presidential Address." *Proceedings of the Aristotelian Society*, 1956, pp. 1–30.
- Balaguer, Mark. *Free Will as an Open Scientific Problem*. MIT P, 2010.
- Barrett, Jeffrey A., and Simon M. Huttegger. "Quantum Randomness and Underdetermination." *Philosophy of Science*, 2020, pp. 391–408.
- Bentham, Jeremy. *An Introduction to the Principles Of Morals and Legislation*. The Online Library of Liberty, A Project Of Liberty Fund, 2011, [oll-resources.s3.us-east-2.amazonaws.com/oll3/store/titles/278/bentham\\_0175\\_EBk\\_v6.0.pdf](https://oll3/store/titles/278/bentham_0175_EBk_v6.0.pdf).
- Bringsjord, Selmer, and Naveen Sundar Govindarajulu. "Artificial Intelligence." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2022, Metaphysics Research Lab, Stanford U, 2022.

- Chalmers, D. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford UP, 1996.
- Churchland, P.M. "Reduction, qualia and the direct introspection of brain states." *Journal of Philosophy*, vol. 82, 1985, pp. 8–28.
- Crane, T. "The mental causation debate." *Proceedings of the Aristotelian Society, Supplementary Volume*, vol. 69, 1995, pp. 211–36.
- Crane, Tim. "The Mind-Body Problem." *The MIT Encyclopedia of the Cognitive Sciences*, edited by Rob Wilson and Frank Keil, Cambridge, MA, USA: MIT P, 1999.
- Curtis, Gary. What are the fallacy files? *What are the Fallacy Files?* [www.fallacyfiles.org/whatarff.html](http://www.fallacyfiles.org/whatarff.html).
- Davidson, Donald. "Mental Events." *Experience and Theory*, edited by L. Foster and J. Swanson, Duckworth, 1970, pp. 79–101.
- . "Thinking causes." *Mental Causation*, edited by J. Heil and A. Mele, Oxford UP, 1993, pp. 1–17.
- Dennett, D.C. *Consciousness Explained*. Harmondsworth: Allen Lane, 1991.
- Descartes, René. *Meditations on First Philosophy*. Caravan Books, 1641.
- Doyle, Arthur Conan. *The Complete Sherlock Holmes*. Doubleday & Co., 1930.
- Eltagouri, Marwa. "A Cubs World Series baby boom? Some parents and hospitals think so." *Chicago Tribune*, 2017. [www.chicagotribune.com/news/breaking/ct-cubs-world-series-baby-boom-met-20170726-story.html](http://www.chicagotribune.com/news/breaking/ct-cubs-world-series-baby-boom-met-20170726-story.html).
- Elzein, Nadine. "Frankfurt-Style Counterexamples and the Importance of Alternative Possibilities." *Acta Anal*, 2017, pp. 169–91.
- Enoch, David. "Why I Am an Objectivist About Ethics (And Why You Are, Too)." *The Ethical Life*, 3rd ed. Edited by Russ Shafer Landau, Oxford UP, 2014.



- extrahistory. The uncanny valley - why more realistic characters look less human - extra credits. *YouTube*, May 2012. [www.youtube.com/watch?v=9K1Kd9mZL8g](http://www.youtube.com/watch?v=9K1Kd9mZL8g).
- Feynman, R. P. *In The character of physical law*. The MIT P, 2017.
- Fischer, John M. "The Frankfurt Cases: The Moral of the Stories." *The Philosophical Review*, 2010, pp. 315–36.
- Fodor, J. A. "Methodological solipsism considered as a research strategy in cognitive psychology." *Behavioral and Brain Sciences*, vol. 3, no. 1, 1980, pp. 63–73. <https://doi.org/10.1017/S0140525X00001771>.
- Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy*, 1969, pp. 829–39.
- Franklin, Christopher Evan. "The Problem of Enhanced Control." *Australasian Journal of Philosophy*, vol. 89.4, 2011, pp. 687–706.
- Half a Life*. 1991. Writer Ted Roberts and Peter Allan Fields, directed by Les Landau, Paramount Domestic Television.
- Hayes, Thomas L. "A Biological View." *Commonweal*, vol. 85, Herodotus. *The Histories*. Translated by Aubrey de Sélincourt, Penguin Books, 1988.
- Hick, John. "The Problem of Evil." *The Encyclopedia of Philosophy*, edited by Paul Edwards, New York: Macmillan, 1967, p. 3.
- Horgan, T. "From supervenience to superdupervenience: meeting the demands of a material world." *Mind*, vol. 102, 1993, pp. 555–86.
- Jackson, F. "Epiphenomenal qualia." *Philosophical Quarterly*, vol. 32, 1982, pp. 127–36.
- . "Mental Causation." *Mind*, vol. 105, 1996, pp. 377–413.
- Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated by Mark Gregor, Cambridge UP, 1998.
- . "On a Supposed Right to Lie From Altruistic Motives." *Critique of Practical Reason, and Other Writings in Moral Philosophy*, edited and translated by Lewis White Beck, U of Chicago P, 1949.

- Kim, J. *Supervenience and Mind*. Cambridge UP, 1993.
- Leibniz, Gottfried Wilhelm. "Theodicy." 1966.
- Lewis, David. "An argument for the identity theory." *Journal of Philosophy*, vol. 63, 1966, pp. 17–25.
- . "Mad pain and martian pain." *Philosophical Papers Volume One*, edited by D. Lewis, Oxford UP, 1983, pp. 122–32.
- . "Reduction of mind." *A Companion to the Philosophy of Mind*, edited by S. Guttenplan, Blackwell, 1995, pp. 412–31.
- . "What experience teaches." *Mind and Cognition*, edited by W. Lycan, Blackwell, 1990, pp. 499–519.
- McCarthy, John. "Ascribing Mental Qualities to Machines." *Philosophical Perspectives in Artificial Intelligence*, edited by Martin Ringle, Humanities P, 1979.
- McGinn, C. "Can we solve the mind-body problem?" *Mind*, vol. 98, 1989, pp. 349–66.
- Mill, John Stuart. *Utilitarianism*. Parker, son, / Bourn, 1863.
- Montesinos, Jesus, et al. "Barcelona baby boom: does sporting success affect birth rate?" *BMJ*, vol. 347, 2013. <https://doi.org/10.1136/bmj.f7387>.
- Nagel, T. "What is it like to be a bat?" *Philosophical Review*, vol. 4, 1974, pp. 435–50.
- Nagel, Thomas. "Death." *Noûs*, vol. 4, no. 1, 1970, pp. 73–80. *JSTOR*, [www.jstor.org/stable/2214297](http://www.jstor.org/stable/2214297). Accessed 14 Feb. 2023.
- Newell, Allen. "Physical Symbol Systems\*." *Cognitive Science*, vol. 4, no. 2, 1980, pp. 135–83. <https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog040202> <https://doi.org/https://doi.org/10.1207/s15516709cog040202>.
- Noonan, John. "Abortion and the Catholic Church: A Summary History." *Natural Law Forum*, vol. 12, 1967.
- . "Deciding Who Is Human." *Natural Law Forum*, vol. 13, 1968.

- Plantinga, Alvin. "The Free Will Defense." *Philosophy in America*, edited by Max Black, Ithaca: Cornell UP, 1965, pp. 204–20.
- Popper, Karl. *The Open Universe: an argument for indeterminism*. Taylor & Francis Group, 1982.
- Pruss, Alexander R. "I Was Once a Fetus: An Identity-Based Argument Against Abortion."
- Pylyshyn, Z. W. "Computation and cognition: issues in the foundations of cognitive science." *Behavioral and Brain Sciences*, vol. 3, 1980.
- Rachels, James. "The Challenge of Cultural Relativism." *Exploring Philosophy: An Introductory Anthology*, edited by Steven M. Cahn, Oxford UP, 1907.
- "Return to Omashu." *Avatar: The Last Airbender*, created by Michael Dante DiMartino, Author Elizabeth Welch Ehasz, directed by Ethan Spaulding, episode 3, Nickelodeon Animation Studio, 2006.
- Richmond, Emily. "The Reality of the Philosophers vs. Welders Debate," 14 Nov. 2015, [www.theatlantic.com/education/archive/2015/11/philosopher-vs-welders/415890/](http://www.theatlantic.com/education/archive/2015/11/philosopher-vs-welders/415890/).
- Robinson, H. *Matter and Sense*. Cambridge UP, 1982.
- Rouzati, Nasrin. "Evil and Human Suffering in Islamic Thought—Towards a Mystical Theodicy." *Religions*, vol. 9, no. 2, 2018. <https://doi.org/10.3390/rel9020047>.
- Rowling, J. K. *Harry Potter And the Sorcerer's Stone*. Arthur A. Levine Books, 1998.
- Schank, R. C., and R. P. Abelson. *Scripts, plans, goals, and understanding*. Erlbaum P, 1977.
- Searle, John. "Minds, Brains, and Programs." *Behavioral and Brain Sciences*, 1980, pp. 417–57.
- Sola, Katie. "Sorry, Rubio, But Philosophers Make 78% More Than Welders," 11 Nov. 2015, [www.forbes.com/sites/katiesola/2015/11/11/rubio-welders-philosophers/?sh=2770ff0641b8](http://www.forbes.com/sites/katiesola/2015/11/11/rubio-welders-philosophers/?sh=2770ff0641b8).

- “Sozin’s Comet.” *Avatar: The Last Airbender*, Creator and writer Michael Dante DiMartino, directed by Joaquim Dos Santos, episode 18-21, Nickelodeon Animation Studio, 2008.
- Talks, Tedx. A teaching assistant named Jill Watson; Ashok Goel. *YouTube*, Nov. 2016. [www.youtube.com/watch?v=WbCguICyfTA](http://www.youtube.com/watch?v=WbCguICyfTA).
- TEDEducation. How to speak monkey: The language of cotton-top tamarins - anne savage. *YouTube*, June 2014. [www.youtube.com/watch?v=4Vfn5CV9juI](http://www.youtube.com/watch?v=4Vfn5CV9juI).
- TEDtalksDirector. Could we speak the language of dolphins? — Denise Herzing. *YouTube*, June 2013. [www.youtube.com/watch?v=CQ5dRyyHwfM](http://www.youtube.com/watch?v=CQ5dRyyHwfM).
- “The Awakening.” *Avatar: The Last Airbender*, created by Michael Dante DiMartino, directed by Giancarlo Volpe, Writer Aaron Ehasz, episode 1, Nickelodeon Animation Studio, 2007.
- “The Beach.” *Avatar: The Last Airbender*, created by Michael Dante DiMartino, Author Katie Mattila, directed by Joaquim Dos Santos, episode 5, Nickelodeon Animation Studio, 2007.
- “The Boy in the Iceberg.” *Avatar: The Last Airbender*, Creator and writer Michael Dante DiMartino, directed by Dave Filoni, episode 1, Nickelodeon Animation Studio, 2005.
- “The Cave of Two Lovers.” *Avatar: The Last Airbender*, created by Michael Dante DiMartino, Author Joshua Hamilton, directed by Lauren MacMullan, episode 2, Nickelodeon Animation Studio, 2006.
- “The Great Divide.” *Avatar: The Last Airbender*, created by Michael Dante DiMartino, Author Aaron Ehasz, directed by Giancarlo Volpe, episode 11, Nickelodeon Animation Studio, 2005.
- “The King of Omashu.” *Avatar: The Last Airbender*, created by Michael Dante DiMartino, Author John O’Bryan, directed by Anthony Lioi, episode 5, Nickelodeon Animation Studio, 2005.

- “The Puppetmaster.” *Avatar: The Last Airbender*, created by Michael Dante DiMartino, Author Tim Hedrick, directed by Joaquim Dos Santos, episode 8, Nickelodeon Animation Studio, 2007.
- “The Warriors of Kyoshi.” *Avatar: The Last Airbender*, created by Michael Dante DiMartino, directed by Giancarlo Volpe, Writer Aaron Ehasz, episode 4, Nickelodeon Animation Studio, 2005.
- “The Waterbending Scroll.” *Avatar: The Last Airbender*, created by Michael Dante DiMartino, directed by Anthony Lioi, Writer John O’Bryan, episode 4, Nickelodeon Animation Studio, 2005.
- Thomson, Judith Jarvis. “A Defense of Abortion.” *Philosophy & Public Affairs*, vol. 1, no. 1, 1971, pp. 47–66.
- Tolstoy, Leo. *Anna Karenina*. Oxford UP, 1980.
- Warren, Mary Anne. “On the Moral and Legal Status of Abortion.” *The Monist*, vol. 57, no. 1, 1973, pp. 43–61. <https://doi.org/10.5840/monist197357133>.
- Weinberg, Justin. “Philosophy Majors Make More Money Than Majors in any other Humanities Field,” 3 Jan. 2019, [www.forbes.com/sites/katiesola/2015/11/11/rubio-welders-philosophers/?sh=2770ff0641b8](http://www.forbes.com/sites/katiesola/2015/11/11/rubio-welders-philosophers/?sh=2770ff0641b8).
- Weizenbaum, J. “Eliza - a computer program for the study of natural language communication between man and machine.” *Communication of the Association for Computing Machinery*, vol. 9, 1965, pp. 36–45.
- Widerker, David. “Libertarianism and Frankfurt’s Attack on the Principle of Alternative Possibilities.” *The Philosophical Review*, 1995, pp. 247–61.
- Winograd, T. “A procedural model of language understanding.” *Computer models of thought and language*, edited by R. Schank and K. Colby, Freeman, 1973, pp. 152–89.
- Wisniewski, et al, David. “Free Will Beliefs Are Better Predicted by Dualism than Determinism Beliefs across Different Cultures.” *PLOS ONE*, vol. 14, 2019.