Introduction to the Philosophy of Nature

Contents

Part I
How to Read Philosophy (This part was written by David W. Concepción)

(Warning: Do not use a highlighter when reading this. As you read on, you'll learn why.)

1 Introduction

Even if you are very smart and very literate, as I assume you are, confusion and frustration may occur if you do not read philosophy in the way philosophers expect you to. There is more than one way to read. In this handout, I describe the basics of How to Read Philosophy.

1.1 What to Expect

Reading philosophy is an activity and, like any activity (e.g., playing volleyball), it takes practice to become good at it. As with any attempt to learn a new skill, you will make some mistakes along the way, get frustrated with the fact that you are progressing more slowly than you would like, and need to ask for help. You may become angry with authors because they say things that go against what you were brought up to believe and you may become frustrated because those same authors argue so well that you cannot prove them wrong. It is likely that you will find unfamiliar vocabulary, abstract ideas, complexly organized writing, and unsettling views. I mention this because it is normal to have certain reactions, such as confusion, outrage, and frustration, when first encountering philosophy. Don't confuse these reactions with failure. Many students who have come before you have had the same initial reactions and succeeded, even your professor.

1.2 The Ultimate Goal

Your aim is to develop, or become more confident in, your personal belief system, by building on what you already know about yourself and the world. By evaluating arguments regarding controversial issues, you should learn to take a well-justified stand that you are able to defend. When you read philosophy you should look for arguments, reasons, and conclusions, not facts, plot or character development, to help you reach your goal of evaluating the plausibility of various positions a person might take on some issue.

1.3 Basic Good Reading Behaviors

- Take care of yourself (take breaks, sit where you won't be distracted, give yourself enough time to read well, sit in an uncomfortable chair to avoid dozing off, etc.)
- Interact with the material (talk to your friends and classmates about what you have read, use a dictionary and philosophical encyclopedia while reading, remember you are reading one person's contribution to an ongoing debate, disagree with the author)
- Keep reasonable expectations (you may not understand everything without some effort, you may need to ask for help or clarification).
- Be able to state the author's conclusion and the gist of the argument for that conclusion BEFORE you come to class.
- Evaluate the gist of the author's argument BEFORE class.
- FLAG and TAKE NOTES. (Flagging is explained below)

2 Important Background Information

2.1 Reading for Information versus Reading for Enlightenment

You are familiar with reading for information: You pass your eyes over some words until some information about the world sticks in your head. Reading for enlightenment may be less familiar. When you read for enlightenment you use a text as an opportunity to reflect upon yourself and your beliefs. Part of the reason why reading for enlightenment is not easy is because self-evaluation often results in personal growth and sometimes when we grow, we experience growing pains.

2.2 Problem-Based, Historical or Figure-Based Philosophy Classes

This is a problem-based class. In problem-based classes, students spend most of their time identifying, reflecting upon, and defending their beliefs. This is not a historical or figure-based course. In historical classes, students spend most of their time learning certain themes in the history of philosophy. In figure-based classes, students spend most of their time mastering what certain philosophers think. In problem-based courses like this one, students read relatively short primary and secondary sources. A secondary source is a text that describes what other people have argued. In this class we won't use secondary sources. A primary source is a text where a person actually argues that a certain position is correct. This course only uses primary sources.

So, you are in a problem-based course where you are supposed to read primary sources for enlightenment. But how, exactly, does one read for enlightenment? Well, strong philosophy readers, people who read with care, do three things. As people increase their ability to read philosophy well they gradually become unaware that they do facet one and they combine facets two and three. However, it is a good idea for non-experts to do one thing at a time.

3 A Three-Part Reading Process

3.1 Part One: Stage Setting

1. Pre-Read:
   For a very short time, examine the general features of the article. Look at the title, section headings, footnotes, bibliography, reading questions, and biography of the author. The goal of the pre-read is to get a basic idea of what the article is about. If you know what an article is about, it is easier to make sense of the individual sentences in it. Also, skim the first and last paragraph to see if you can easily identify a focal or thesis statement. A focal statement describes the topic of the text. Focal statements often begin with phrases such as "I will discuss X, Y, and Z." A thesis statement is a more specific description of the author's goal. Thesis statements often begin with phrases such as "I will show that X is true and Y and Z are false." While doing the pre-read, ask yourself "How am I doing?" by answering the following questions:

   (a) Is this a primary or secondary text?
   (b) Should I expect an argument or a description of an argument?
   (c) Am I reading for information or enlightenment?
   (d) What is the focal statement of the article?
   (e) Is there a thesis statement? What is it?
   (f) What should I expect to find in the text in light of the title?
   (g) Are there section headings? If yes, what can I learn about the article from them?
   (h) Is there a bibliography? If yes, what can I learn about the article from it?
   (i) Are there footnotes? Are they essentially documentation or do they say something? (This lets you know whether you need to read them when you see a number in the text.)

2. Fast-Read:
   Read the entire article fairly quickly. The goal of the fast-read is to develop a basic understanding of the text. When doing the fast read, remember to do the following:

   (a) Identify the thesis statement. Warning: You may not be able to do this until you reach the end of the article. Mark anything that seems like it might be a thesis statement or conclusion when you first notice it, then pick the one that seems most central when you are done. In some cases, the author may not even actually write a thesis statement down, so you may need to write one for the author.

   (b) Look up definitions of words you don't know and write them in the margins. Warning: Don't get bogged down while doing this. If it is too difficult to figure out which meaning of a term an author seems to have in mind, or if you have to read an entire encyclopedia entry to figure out the meaning, just move on.
   (If you read near a computer see The Free Dictionary: `http://www.thefreedictionary.com/` and Stanford Encyclopedia of Philosophy: `http://plato.stanford.edu/`.)

   (c) FLAG the structure of the article in as much detail as possible without getting bogged down. When you flag a text you put marks in it that will allow you to reconstruct the meaning of the text without having to reread the entire text again. See below for specific suggestions on how to flag an article.

   (d) Don't let anything stop your progress. This is a fast read. You may skim long examples.

   (e) Ask yourself "How am I doing?" by answering the following questions:

      i. Have I identified the thesis statement and written it down?

      ii. Do I know what the conclusion of the author's argument is and have I marked places in the text where important steps toward that conclusion occur?

3.2 Part Two: Read for Understanding

Develop a sophisticated understanding of the text. You should be able to explain to a friend how the author defends her/his conclusion. Once you are able to coherently explain the article in your own words, you have truly internalized it—good job!

When reading for understanding, remember to do the following:

1. Re-read the entire article VERY CAREFULLY.

2. Correct and add to your previous flagging.

3. Take lots of notes. In these notes, rephrase what the author says in your own words. Remember: You should practice the principle of charity when taking notes. Describe the author's view in the most favorable way possible. If you have trouble taking notes, stop at the end of every section or paragraph (sometimes even every sentence) and mentally rephrase the meaning of the text in your own words.

4. Draw diagrams or flow charts of the major moves in the article if doing so helps you.

5. Bring together all your work so far into a summary that is detailed enough that you won't have re-read the article again to remind yourself of the author's argument.

6. While reading for understanding, ask yourself "How am I doing?" by answering the following questions:

   (a) Do I know exactly what the author is saying?

   (b) Have I re-re-read passages that were confusing at first?

   (c) Can I connect the dots?

   (d) Can I explain in my own words why the author concludes what she or he concludes? (In the fast-read you find the conclusion and do your best to figure out the steps to it. In the read-for-understanding, you come to fully understand each step in detail.)

3.3 Part Three: Evaluating

Now that you have made yourself a concise and easy to articulate summary of the author's argument, it is time to evaluate it. When evaluating, your main tool is the summary you made, but you will need to re-re-read certain passages. At this stage, you are entering the debate, rather than simply learning about it.

When evaluating a text, remember to do the following:

1. Fix any mistaken flagging as you re-re-read important passages.
2. Write down anything new that you discover as you go through the text again.
3. While evaluating a text, ask yourself "How am I doing?" by answering the following questions:
   (a) Have I looked to see if every conclusion in the text is well defended?
   (b) Have I thought about how an undefended conclusion could be defended? (Have I been charitable?)
   (c) Do I think the arguments for the conclusions are persuasive? Why or why not?
   (d) Can I think of any counter-examples to any assertion made by the author?
   (e) Can I put my finger on exactly what bothers me about what the author says?
   (f) Can I explain where and why I think the author made a mistake? Have I thought about how the author might respond to my criticism?
   (g) Have I identified some of my own beliefs that can't be true if the author is right?
   (h) Is there is a conflict between what I believe and what the author says? If so, to avoid being a hypocrite I must ultimately change my mind or show that the author's reasoning fails in some way. Simply identifying a disagreement does not constitute an evaluation. Have I figured out, exactly, what the author got wrong so that I may continue to believe as I always have with confidence? Have I figured out, exactly, which of my beliefs I must change in light of what I have learned from the author? Have I looked for some point that the author did not consider that might influence what I think is true?

4 How do I FLAG?

When you flag a text you put short notes, preferably *in pencil*, in the margins of the text (unless you are using a library book) that will remind you of many details in the text so that you will not have to reread an entire text to reconstruct its meaning in your head. Flagging marks allow you to pick out various important features of the text for further study. Flagging is better than highlighting because flagging is more detailed than highlighting. If all you're interested in doing is distinguishing something that seems important from other stuff that doesn't seem important then highlighting is fine. But you want to do more than just distinguish important from unimportant. There is more than one kind of important thing in a philosophy text, and you want to mark your text in such a way that you can tell the difference. Another good thing about flagging is that you can "unflag" and you can't "unhighlight. " The flexibility to change your notes is important because sometimes as you read further into a text, or read it a second time, you realize that something that seemed important really isn't important.

There are many ways to flag a text. You should develop your own method and notations. You should use whatever marks help you attain the goal of noting the different types of important parts of a text. A part of the text is important when it must be present for the author's conclusion to make sense. On some occasions important things are a sentence or a clause in length, but other times important things are a paragraph or a page long. The following are suggestions of abbreviations that have been particularly useful to me. But, again, feel free to use terms not on the list that you find helpful and ignore any, or all, of these if you find them unhelpful. In addition to these terms, I circle "list" words (e.g., First, second, [i], [ii]) and I underline definitions.

4.1 Tracking the Flow

Focal General topic this article will discuss

Thesis Specific claim the author hopes to prove

Dfn Definition

Dst Distinction

e.g. Example

CI Important claim the author will argue is true, or an important assertion of fact

Arg Argument or reason to believe an assertion or conclusion; justification of a claim

Obj Objection

Reply Reply to an objection

Sum Summary

4.2 Self Monitoring

??? What? I don't get it. I must reread this passage carefully

=x? This means what exactly?

4.3 Reader Evaluation

Why? Why should someone agree with this?

[Underline] This is important!

Flagging should naturally evolve into note taking. If you are inclined to write "???" in a margin, it is a good idea to write out more fully what confused you. If you can articulate your confusion you are a good way down the road to figuring out what's going on. During your reread for understanding, make sure to spend as much time as necessary to fully grasp what is going on in the "???" section.

4.4 Key Words

Some students find the following list of words or phrases that signal a significant moment in a text helpful. However, there are many texts where authors will not use any of these terms or phrases. These are words or phrases to be aware of so that if they come up you are ready, but you should not read a text as if you are on a treasure hunt for these words or phrases.

1. Focal statements are often signaled by phrases such as: I will discuss, Consideration will be given to, My main concern is
2. Thesis statements are often signaled by such as phrases such as: In this paper I argue that, I hope to conclude that, I will show that
3. Premises, Reasons, or Assertions are often signaled by words or phrases such as: Because, Since, For, Whereas, Secondly, Given that, As shown or indicated by, The reason is that
4. Objections or criticisms are often signaled by words or phrases such as: Moreover, However, It could be objected that, Opponents of my view might claim, Critics might say, On the other hand, There is reason to doubt

5. Replies or Rejoinders are often signaled by words or phrases such as: This criticism fails because, My opponent does not notice that, In response we should remember, Nevertheless, On the other hand

6. Conclusions are often signaled by words or phrases such as: In summary, Thus, Therefore, So, Hence, Accordingly, Consequently, As a result, We may infer, Which entails that, It follows that

5 A Final Complication

5.1 Linear versus Dialogical Writing

Students sometimes ask me one or all of the following questions: (1) Why does the author contradict herself? (2) Why does the author repeat himself so much? (3) Why is this reading so wordy? Students ask these questions, I think, because they expect the reading to be linear when, in fact, philosophical writing is usually dialogical. So, let me tell you a little bit about dialogical writing and then I will answer each question individually. Linear writing moves in a straightforward way from one idea to the next, without examining (m)any supporting or contradictory ideas. Dialogical writing explicitly acknowledges and responds to criticism. It may be helpful to think of philosophical writing as a monologue that contains a dialogue. The author is speaking directly to you, delivering a monologue for your consideration. But in the monologue, the author is telling you about a dialogue or debate that she or he knows about, while giving you reasons for thinking that her or his understanding of that debate is right. As you know, in some debates there are more than two sides and sometimes people on the same side have different reasons for believing what they believe. Authors will take the time to tell you about as many sides, or different camps within one side, as they think you need to know about to understand, and be persuaded by, their view. This confuses people sometimes because it is hard to keep track of whether the author is arguing for their side or talking from another point of view or camp within the same side for the sake of (good) argument. Points to remember about dialogical philosophical texts:

1. Authors sometimes support their views with thought-experiments (i.e., examples that ask you to imagine how things would be if something that is not true, were true).

2. Authors sometimes argue that other thinkers haven't noticed an important difference between two things. Authors draw distinctions.

3. Authors sometimes argue that another philosopher's views or arguments ought to be rejected. There is something really tricky here. Fair-minded writers will practice the principle of charity. According to the principle of charity, one should give one's opponents the benefit of the doubt; one should respond to the best thing that someone who disagrees with you could say, even if they didn't notice it. Sometimes attempts to abide by the principle of charity results in authors presenting arguments for the correctness of views they ultimately reject. That is, for the sake of (good) argument some authors will present reasons for thinking that their critics are right. Try to avoid mistaking charitable elucidation for the author's main argument.

Now that you are more familiar with dialogical texts I can answer the questions students sometimes ask about them.

5.2 Frequently Asked Questions

1. Why does the author contradict herself?
   Sometimes thinkers do unwittingly contradict themselves. Most of the time, however, people perceive a contradiction where there isn't one because they fail to notice a change in "voice." Authors will describe many sides, and camps within a side, but they will voice agreement with only one side or camp. If you lose track of the fact that the author is considering an alternative view, you will mistakenly think that a fair-minded examination of a different point of view is a contradiction. Keeping track of where you are in the argument is crucial to understanding. If you think you see a contradiction, double or triple check your flagging to make sure that you are not simply missing something.

2. Why does the author repeat himself so much?
   Usually philosophers do not repeat themselves all that much. Sometimes, however, they use examples that are so long, or discuss material that is interesting but ultimately tangential for such a long time, that they (correctly) assume that their readers have lost track of the point being made. In such cases, a simple repetition may occur for the benefit of the reader. More often, however, people lose track of where they are in an argument and consequently mistake something new for repetition. Again, keeping track of where you are in an argument is crucial to understanding and flagging really helps readers keep track of where they are.

3. Why is the writing so wordy?
   Some people think philosophers use all sorts of fancy words to intimidate their readers or show off. This reaction is understandable but mistaken in at least three ways:
   First, it is a mistake to become angry with an author because you have a limited vocabulary. There is an opportunity for learning here. Take it!
   Second, there is an international community of philosophers, and like all specialized communities (such as you and your friends), there are certain patterns in the way members of that community talk to one another. Metaphorically, when you enrolled in philosophy class you walked into a room where a bunch of people have been having a conversation for a very long time. You need to adapt to their idiosyncratic ways of talking if you want to participate in their conversation. Of course, philosophers shouldn't be rude and intentionally try to exclude you with their words. But it is important to realize that they didn't know you were coming, so they might not have done everything possible to make your inclusion as easy as you would like. Whatever the author's faults, do your part - be open to what is being said, try your hardest to understand, and don't assume the worst about the author, even if the author doesn't always behave as you would like.
   Third, and most importantly, not every complex idea can be stated in simple terms. Sometimes simplification is over-simplification, where the important nuances of what a person really thinks are lost. It is true that some philosophical writing is more complicated than it needs to be, but not all of it is. Some philosophical writing needs to be complicated to express a complicated idea. Part of the beauty of philosophy is its complexity. Do your best to appreciate the beauty of complexity.

6 Summary: What Successful Philosophy Readers Do

- Abide by the "Basic Good Reading Behaviors".
- Before class, complete all three parts of the Three-Part Reading Process.
- Flag and Take Notes to keep track of where you are in the dialogue.

Part II
Are We Merely Physical Systems? Do We Have Souls?

Texts:

- Gertler, B. (2007). In Defense of Mind-Body Dualism. In Feinberg, J. and Shafer-Landau, R. (eds.), Reason and Responsibility, 13th edition. Wadsworth, pp. 312-319. [Excerpt; read only sections 1-4 (inclusive)]

7 Physicalism and Dualism

[Physicalism/Materialism] All the facts (= instantiations of properties or relations) obtain in virtue of the physical facts (= instantiations of physical properties or relations).

This entails that:

[Supervenience] All the facts supervene on the physical facts, i.e.,

It is impossible that the physical facts are the same as the actual physical facts, but some other fact is not the same as an actual fact, i.e.,

Every minimal physical duplicate of our world is a duplicate of our world in every respect.

(Here, by "physical property or relation" I mean a property or relation invoked by the completed theory of physics.)

Standard initial motivations for physicalism come from

- Disbelief in the unnatural, and the success of modern science.
- The reductive program in the sciences

[Dualism] Some facts (which most dualists say are mental facts) do not supervene on the physical facts, i.e.,

It is possible that the physical facts are the same as the actual physical facts, but some other fact (which most dualists say is a mental fact) is not the same as an actual fact, i.e.,

Some minimal physical duplicate of our world is not a duplicate of our world in every respect (and most dualists say it differs in a mental respect).

Standard initial motivations for dualism come from

- The fact that our mental states can be about things outside our heads.
- The "hard problem" of phenomenal consciousness: The problem of explaining how physical systems give rise to phenomenally conscious experience, i.e., to states such that there is something it is like to be in them.

Exercise Below are a few claims. Which of them entails that physicalism is false?

1. Alice could have had three eyes.
2. Bob could have had three eyes, and green things could have looked different to him than they actually do (e.g., they could have looked red to him).
3. All the physical facts could have been just as they actually are while green things looked different to Bob than they actually do.
4. All the physical facts could have been just as they actually are while all the people are zombies - there is nothing it is like to be them.
5. A non-physical angel exists.
6. A non-physical soul could have existed.
7. All the mental facts could have been just as they actually are while there are no physical bodies.

8 Identity Physicalism

[Identity Physicalism] Every mental property is identical to some physical, chemical or biological property.

(E.g., being pain = being C-fiber stimulation, or being pain = being basal ganglia stimulation, or ...)

Exercise Discuss the following questions:

1. Does this entail [Physicalism]? Why?
2. What is the difference between identity and correlation? For instance, what is the difference between saying that having a heart is identical to having kidneys and saying that something has a heart if and only if it has kidneys?
3. Is everything necessarily self-identical? Is Superman necessarily identical to Clark Kent? Is water necessarily identical to H2O?
4. What does the previous question (#3) have to do with identity physicalism?
5. What is the difference between conceivability and possibility?

9 The Disembodiment Argument

Version A

1. I can conceive of experiencing this very pain while possessing no physical features. In other words, I can conceive of experiencing this very pain while disembodied.

2. If I can conceive of a particular scenario occurring, then that scenario is possible.

So

3. It is possible that this very pain occurs in a disembodied being.

4. If this very pain were identical to some physical state, then it could not possibly occur in a disembodied being.

So

5. This very pain is not identical to any physical state.

So

Conclusion. [Identity Physicalism] is false.

Exercise

1. Is the disembodiment argument effective against all versions of physicalism, including those that reject identity physicalism?
2. What is wrong with premise 2?

Version B

1*. Using concepts that are sufficiently comprehensive I can conceive of experiencing this very pain while disembodied.

2*. If, using concepts that are sufficiently comprehensive, I can conceive of a particular scenario occurring, then that scenario is possible.

So

3. It is possible that this very pain occurs in a disembodied being.

4. If this very pain was identical to some physical state, then it could not possibly occur in a disembodied being.

So

5. This very pain is not identical to any physical state.

So

Conclusion. [Identity Physicalism] is false.

In defense of premise 1*:

6. Being pain = feeling like pain, i.e., hurting. (This implies that our concept of "pain" is a concept of a mental state with no hidden essence.)

7. Using concepts that are sufficiently comprehensive I can conceive of something hurting me in this very way while disembodied.

So

1*. Using concepts that are sufficiently comprehensive I can conceive of experiencing this very pain while disembodied.

Part III
What Is It To Have Toothache, To Taste An Apple's Flavor, Or To Experience The Color of the Sky?

Texts:

- Putnam, H. (2002). The Nature of Mental States. In Chalmers, D. J. (ed.), Philosophy of Mind: Classical and Contemporary Readings. Oxford: Oxford University Press, pp. 75-79. [Excerpt; read only sections 2-5 (inclusive)]

10 Functionalism

[Functionalism] Every mental property is identical to some *functional* property.

> **[Machine Functionalism]** Every mental property is identical to some property of the kind *being in state S of probabilistic automaton M*, where a probabilistic automaton is (for present purposes) a machine which
>
> 1. at any time is in one of finitely many states, can receive one of finitely many sensory inputs and produce one of finitely many sensory outputs, and
> 2. for any state the machine can be in, and for any sensory input the machine can receive, there is some fixed probability that the machine will transition from its state and received input to another one of its states and produce some behavioral output.
>
> **[Psycho-functionalism]** Every mental property is identical to some property of the kind *being disposed (i) to be caused by such and such sensory inputs and by instantiations of such and such mental properties, and (ii) to cause such and such behavioral outputs and instantiations of such and such mental properties*.
>
> (E.g., being pain = being disposed (i) to be caused by stimulation of such and such receptors—so-called "pain receptors"—and (ii) to cause fear, as well as fight or flight behavioral outputs.)

Since a non-physical angel could, in principle, have functional properties, functionalism is strictly speaking compatible with dualism. However, most functionalists believe that all the mental facts (which they take to be instantiations of functional properties and relations) obtain in virtue of the physical facts (= instantiations of physical properties or relations). For this reason, functionalism is often seen as a version of physicalism about the mental.

Motivations for functionalism:

- No single mental state can be identified with a tendency to behave in certain ways. This is because, first, the way we tend to behave depends on more than a single mental states; and second, our mental states are the things that *explain* our behavioral tendencies rather than our behavioral tendencies themselves. Functionalism explicitly acknowledges both these points.
- If mental properties are identified with functional properties, then the laws of psychology may well be derivable from the laws governing the interactions between an organism's internal states, its sensory inputs, and its behavioral outputs. This promotes the reductive program in science.
- If mental properties are identified with functional properties, then there is no more question about what the nature of mental properties is. It is simply the nature of functional properties.

Note. The view I have described simply as functionalism is sometimes called "role functionalism". There is a related view, sometimes called "realizer functionalism", which I will not discuss. This latter view is somewhere between role functionalism and identity physicalism.

11 Functionalism vs. Identity Physicalism

The Multiple Realizability Argument :

1. Pain (like other mental properties) is *multiply realizable*: A mammal, a reptile, a mollusk (e.g., an octopus) and perhaps aliens from other planets can all feel pain; but other organisms cannot feel pain.

2. Therefore, if being pain is identical to some physical, chemical or biological property, then there is a single physical, chemical or biological property that mammals, reptiles, mollusks and perhaps aliens can all have, but that organisms that cannot feel pain cannot have.

3. It is implausible that there is a single physical, chemical or biological property that mammals, reptiles, mollusks and perhaps aliens can all have, but that organisms that cannot feel pain cannot have.

4. Therefore, being pain is *not* identical to some physical, chemical or biological property.

Exercise

1. Is functionalism itself vulnerable to something like the multiple realizability argument? Why?
2. Why is the disembodiment argument ineffective against functionalism?
3. Can you replace the the disembodiment argument with a similar argument that is effective against functionalism?

Part IV
How Can Decisions Lift Us From Our Chairs? And How Can An Empty Stomach Cause Us To Feel Hungry?

Texts:

- Kim, J. (2005). The Supervenience Argument Motivated, Clarified and Defended. In his Physicalism, Or Something Near Enough. Prince-ton, NJ: Princeton University Press, pp. 32-45. (Excerpt.)

12 Mental Causation

- Mental to physical causation. For example: Deciding to get out of bed and go to work causes me to get out of bed and go to work.
- Physical to mental causation. For example: Having an empty stomach causes me to feel hungry.

Cases of mental to physical causation, and even cases of mental to mental causation (e.g., deciding to get out of bed and go to work causes me to think about the road I need to take to get to work) have been thought to pose a serious problem for dualism and non-reductive physicalism.

13 Non Reductive Physicalism

[Physicalism/Materialism] All the facts obtain in virtue of the physical facts.

This entails [Supervenience]: All the facts supervene on the physical facts, i.e., it is impossible that the physical facts are the same as the actual physical facts, but some other fact is not the same as an actual fact.

[Physicalism About the Mental] All mental facts obtain in virtue of the physical facts.

This entails [Supervenience About the Mental]: All the mental facts supervene on the physical facts, i.e., it is impossible that the physical facts are the same as the actual physical facts, but some mental fact is not the same as an actual fact.

[Identity Physicalism] Every mental property is identical to some physical, chemical or biological property.

[Non Reductive Physicalism] = [Physicalism About the Mental] is true, but *no* mental property is identical to some physical, chemical or biological property.

Most, perhaps all, functionalists are non-reductive physicalists.

Exercise: Why?

14 Kim's Causal Exclusion Argument (The Supervenience Argument)

14.1 Three Principles

1. [Causal Closure of the Physical] If the instantiation of a physical property is caused at a given time by the instantiation of some property, it is caused by the instantiation of a physical property at that time.

The [Causal Closure of the Physical] principle is often accepted on the grounds of the success of contemporary physical science, and specifically on the grounds of empirical evidence in favor of energy, mass and momentum conservation laws.

2. [Edward's Dictum] If an instantiation of property X both

(i) supervenes on an instantiation of property Y (i.e., Y is instantiated, and it is impossible for Y to be instantiated when X is not instantiated.), and

(ii) is caused by an instantiation of property Z,

then the instantiation of property Z caused the instantiation of property Y.

3. [Exclusion] No single instantiation of a property can be (sufficiently) caused by more than one instantiation of a property at a time.

14.2 The Argument

With these three principles we can formulate the following argument against [Non Reductive Physicalism]:

1.	An instantiation of mental property M at t causes an instantiation of mental property M*.	Hypothesis
2.	There is some physical property P* such that M*'s instantiation supervenes on P*'s instantiation. (In other words: There is some physical property P* such that P* is instantiated, and it is impossible for it to be instantiated when M* is not instantiated.)	From 1, by [Non Reductive Physicalism]
3.	The instantiation of mental property M at t causes the instantiation of physical property P*	From 1 and 2, by [Edward's Dictum]
4.	The instantiation of physical property P* is caused by the instantiation of some property at t.	From 3
5.	The instantiation of physical property P* is caused by the instantiation of some physical property P at t.	From 4, by [Causal Closure of the Physical]
6.	Physical property P and mental property M are distinct.	By [Non Reductive Physicalism]
7.	The instantiation of P* is caused by more than one instantiation of a property at a time.	By 3, 5, and 6
8.	No single instantiation of a property can be (sufficiently) cause by more than one instantiation of a property at a time.	[Exclusion]
9.	Contradiction!	By 7, and 8

The contradiction shows that [Edward's Dictum], [Causal Closure of the Physical], and [Exclusion] are incompatible with the combination of [Non Reductive Physicalism] with our Hypothesis. It also shows that [Edward's Dictum], [Causal Closure of the Physical], and [Exclusion] are incompatible with the combination of [Non Reductive Physicalism] with line 3. Therefore, if you accept [Edward's Dictum], [Causal Closure of the Physical], [Exclusion] and [Non Reductive Physicalism], you have to deny that instantiations of mental properties can cause either instantiations of other mental properties or instantiations of other physical properties. In short, you are forced to accept

[Epiphenominalism] Instantiations of mental properties cause neither instantiations of mental properties nor instantiations of physical properties.

Exercise: Why is this argument a problem for a dualist?

14.3 A Possible Reply?

Non reductive physicalists have denied [Exclusion]. They suggest that a noise at my door can be caused both by both a knock on my door at t and by a strong knock on my door at t. This think this can happen, even though being a knock is distinct from being a strong knock, when the knock supervenes on the strong knock. Similarly, they suggest, something can be caused both by the instantiation of a physical property at t and by the instantiation of a mental property at t, when the mental property supervenes on the physical property.

Part V
Can We Know Anything About The External World?

Texts:

- Stroud, B. (2000). The Problem of the External World. In Sosa, E., Kim, J. and McGrath, M. (eds.), Epistemology: An Anthology. Malden, MA: Blackwell, pp. 6-23.
- Moore, G. E. (2000). Proof of an External Worlds. In Sosa, E., Kim, J. and McGrath, M. (eds.), Epistemology: An Anthology. Malden, MA: Blackwell, pp. 24-26.

15 Cartesian Skepticism

Consider *The Matrix* scenario:

> You believe lots of things about what the world outside you is like, and about how you are situated in that world. For example, you believe (i) that you are a member of Ben Gurion University, (ii) that you sometimes meet friends and family members in person, (iii) that you have some money in the bank, (iv) that you occassionally spend money on food, drink, shelter and other things, and (v) that you have a name that was given to you when you were young. However, although you believe all these things to be the case, they are not the case. In fact, the external world is a post-apocalyptic waseland, in which you and all other human beings are agriculturally farmed in a field. To keep the humans docile, every human (including you!) is intravenously fed and kept connected to a computer simulation, called "The Matrix", which is sending signals to the human's sensory nerves. These signals causae the humans to have hallucinations that keep them unaware of their real condition, and so docile. In your case, the signals you receive from The Matrix cause you to have the experiences which mislead you into having the external world beliefs that you do.

Now, when we evaluate this scenario, it seems like we can take the following for granted:

(The Matrix is Bad) That *The Matrix* scenario is real enatils that you do *not* know a great many of the things you believe.

However, it is not obvious that *The Matrix* scenario *is* real. So it is not obvious that you do *not* know a great many of the things you believe.

Nevertheless, Cartesian skeptics think that you do not know a great many of the things you believe. And they think that they can *show* that you do not know those things, even *without* showing that *The Matrix* scenario is real! Here is their argument:

15.1 Cartesian Skepic's Argument

(Seemings Don't Exclude The Matrix) The ways that things seem to you to be do not guarantee that *The Matrix* scenario is not real. (After all, if it were real, things would seem to you to be just as they do seem to you to be.)

(Non-Excluded Matrix leads to Matrix-Ignorance) If (Seemings Don't Exclude The Matrix)—i.e., if the ways that things seem to you to be do not guarantee that *The Matrix* scenario is not real—then you do not know that *The Matrix* scenario is not real.

(Matrix-Ignorance) Therefore, you do not know that *The Matrix* scenario is not real.

(Ignorance Inflation) If you do not know that *The Matrix* scenario is not real, then you do not know a great many of the things you believe.

(Skepticism) Therefore, you do not know a great many of the things you believe.

On the one hand, the argument is plausible. On the other hand, its conclusion is difficult to accept. Consider Moore's argument:

15.2 Moore's Argument

(Hands) I can argue that two hands exist using the following "hand-argument":

- This is a hand.
- This is another hand.
- Therefore, two hands exist.

(Known Premises) I know the premises of the hand-argument.

(Known Conclusion) Therefore, I know the conclusion of the hand-argument, i.e., I know that two hands exist.

(Non-skepticism) Therefore, I know a great many of the things I believe.

Moore thinks this counterargument refutes the Cartesian skeptic. Of course, a Cartesian skeptic would reply that since her own argument for skepticism is sound, Moore's argument is unsound. The Cartesian skeptic would likely add that the reason Moore'sd argument is unsound is that (Known Premises) is false, i.e., that I do not know the premises of the "hand-argument". We therefore have a stalemate: Moore says his own argument shows that the Cartesian skeptic's argument is unsound, while the Cartesian skeptic says that her own argument shows that Moore's argument is unsound.

We need to settle the stalemate. One way is to scrutinize Moore's argument. The other way, and the way we will pursue today, is to scrutinize the skeptic's argument. More specifically, we will take a close look at just two the the premises the skepic is relying on: The (Non-Excluded Matrix leads to Matrix-Ignorance) premise, and the (Ignorance Inflation) premise.

16 The (Non-Excluded Matrix leads to Matrix-Ignorance) Premise

The premise is:

(Non-Excluded Matrix leads to Matrix-Ignorance) If the ways that things seem to you to be do not guarantee that *The Matrix* scenario is not real, then you do not know that *The Matrix* scenario is not real.

Why would anyone believe this? One reason to believe this comes from considerations regarding *certainty*. To bring these consideratins out, consider the *Lottery* scenario:

> You have decided to play in a lottery, and purchased exactly one lottery ticket. Your chance of winning the lottery is 1/3,000,000. Your chances of losing are therefore 2,999,999/300,000.

Intuitively, you don't *know* that you will lose the lotter. Why? Well,

- First, if you *do* know that you will lose, you probably knew that you will lose even before you bought the ticket. But then why did you buy the ticket to begin with?
- Second, there is nothing special about your ticket, compared to all the others. So, if you know that *you* will lose, you probably also know that every other ticket will lose too. But you can't know that, since you know that some ticket will win!

What is Lottery scenario seems to suggest is that even when you have *very* good reasons to believe something, as long as there is a small chance that you might be wrong, you do not know it. In other words:

(Inconclusiveness) If your basis for believing that p is not conclusive (i.e., if it does not guarantee that p), then you do not know that p.

Now, you might also think that

(Seemings as Basis) If the ways that things seem to you to be do not guarantee that *p*, then your basis for believing that *p* is not conclusive (i.e., it does not guarantee that *p*).

Together, (Inconclusiveness) and (Seemings and Basis) entail that

(Non-Excluded Alternatives leads to Ignorance) If the ways that things seem to you to be do not guarantee that *p*, then you do not know that *p*.

The (Non-Excluded Matrix leads to Matrix-Ignorance) premise is just an instance of this general principle. Importantly, however, both (Inconclusiveness) and (Seemings as basic) are disputed.

- (Inconclusiveness) is disputed by those who claim that we often *do* know things on the basis of inconclusive reasons. For example, they think we can know that our car is out of petrol by reading the dashboard petrol meter, even though there is always some chance that the meter is broken. Defenders of (Inconclusiveness) reply by saying that if your basis for believing that *p* is not conclusive, then for all you have basis to believe it is not the case that *p*. But, if for all you have basis to believe it is not the case that *p*, then you do not know that *p*.
- (Seemings as Basis) is disputed by those who claim that your basis for believing something can be something other than the way things seem to you to be. On some of these views, your basis for believing something can be such that it is not a mental state of yours, and can even be such that you cannot know you have that basis. For example, on some views small childred (who have never learnt anything about grammar) know which sentences in their mother tongue are grammatical, without ever having considered whether those sentences are grammatical (and so, without those sentences ever seeming grammatical to them), without the basis for their knowledge being any mental state they are in, and without their being able to know that they have a basis for their knowledge.

17 The (Ignorance Inflation) Premise

The premise is:

(Ignorance Inflation) If you do not know that *The Matrix* scenario is not real, then you do not know a great many of the things you believe.

Why would anyone believe this? We will look at two reasons to believe (Ignorance Inflation). The first has to do with considerations regarding *knowledge and logical entailment*, and the second has to do with considerations regarking *higher-order knowledge*. We'll take them in turn.

17.1 Knowledge and Entailment

We we have both a set of things, and a 2-place relation that members of the set can stand in, we can ask whether the set *is closed under* the relation. We say that a set *S* is closed under a relation *R* if and only if whenever one thing is both a member of *S* are bears relation *R* to another thing, then the other thing is also a member of set *S*. For example:

- The set of *colored things* is closed under the relation *less red than* (because no matter what *x* and *y* are, whenever *x* is both colored and less red than *y*, *y* is also colored.)
- The set of *shaped things* is closed under the relation *has a smaller area than (*because no matter what *x* and *y* are, whenever *x* is both shaped and has a smaller area than *y*, *y* is also shaped.)
- The set of *my ancestors* is closed under the relation *son of* (because no matter what *x* and *y* are, whenever *x* is both an ancestor of mine and the son of *y*, *y* is also an ancestor of mine.)

Why is this important? Well, because Stroud (e.g.) thinks that in odinary life, we behave as though *knowledge is closed under entailment*. Strictly speaking, to say that knowledge is closed under entailment is to say that no matter what x and y are, whenever x is both something that S knows to be the case and that entails y, y is also something that S knows to be the case. However, in epistemology it is common to use the name "knowledge closure principle" not just for the last claim, but for any of the following claims (and for others like them):

(Knowledge is Closed Under Entailment) If both (i) that p entails that q and (ii) S knows that p, then S knows that q.

(Knowledge is Closed Under Known Entailment) If (i) S knows that p entails that q and (ii) S knows that p, then S knows that q.

(Knowledge is Closed Under Competent Deduction) If both (i) S can competently deduce that q from the premise that p and (ii) S knows that p, then S is in a position to know that q.

Now, if any of these closure claims are true, we can use them in an argument for (Ignorance Inflation). To simplify things, we will work with a version of this argument that uses the simple (Knowledge is closed under entailment) claim. The argument goes like this:

(Definition of The Matrix Scenario) That you am a member of Ben Gurion University entails that *The Matrix* scenario is not real.

(Since Knowledge Is Closed Under Entailment) If both (i) that you are a member of Ben Gurion University entails that *The Matrix* scenario is not real and (ii) you know that you are a member of Ben Gurion University, then you know that *The Matrix* scenario is not real.

(Excluded Matrix) Therefore, if you know that you are a member of Ben Gurion University, then you know that *The Matrix* scenario is not real.

(Ignorance Inflation*) Therefore, if you do *not* know that *The Matrix* scenario is not real, then you do *not* know that you are a member of Ben Gurion University.

Notice that we could run this argument not just with the claim that you are a member of Ben Gurion University, but also with the calim that you have a name that was given to you when you were young, or with the claim that you sometimes meet friends and family in person, etc. In fact, we could run this argument with any of the great many claims that you believe and which entails that *The Matrix* scenario is not real. For this reason, we can deduce from the last argument that

(Ignorance Inflation) If you do not know that *The Matrix* scenario is not real, then you do not know a great many of the things you believe.

17.2 Higher-Order Knowledge

The arguement for (Ignorance Inflation) which we just saw relied on the particular way in which we defined The Matrix scenario. Specifically, it relied on the fact that we defined the scenario in such a way that the following is true: That you are a member of Ben Gurion University entails that *The Matrix* scenario is not real. More broadly, we relied on the fact that we defined the scenario in such a way that its reality entailed the *falsity* a great many of your beliefs.

One of the most intriguing things about (Ignorance Inflation), however, is that it is possible to argue for it even *without* the assumption that the reality of *The Matrix* scenario entails the falsity of many of your beliefs! In fact, we can redefine *The Matrix* scenario so as to be compatible with your beliefs' being *true*, and still argue for (Ignorance Inflation). To see how this might be so, let's start by re-defining *The Matrix* scenario thus:

> You believe lots of things about what the world outside you is like, and about how you are situated in that world. For example, you believe (i) that you are a member of Ben Gurion University, (ii) that you sometimes meet friends and family members in person, (iii) that you have some money

in the bank, (iv) that you occassionally spend money on food, drink, shelter and other things, and (v) that you have a name that was given to you when you were young. However, you only believe these things because some machine is causing you to have experiences which in turn cause you to believe the things that you do.

In many ways, this revised scenario is similar to the original one. In this scenario too, it seems clearly right that:

(The Matrix is Bad) That *The Matrix* scenario is real enatils that you do *not* know a great many of the things you believe.

Furthermore, since it is not obvious that the revised scenario *is* real, it is also not obvious that you do *not* know a great many of the things you believe. In all these ways the scenarious are similar.

Here, however, is how the sceanrios differ: The revised scenario could be real, and still your beliefs about the external world could be mostely true. Why? Well, the scenario only says that your beliefs and experiences are not related to reality in the right way. But the fact that they arte not realted to reality in the right way does not mean that reality is not, by some accident, just as you believe it to be. Therefore, the scenario can be real, and your beliefs about the external world can fail to be related to reality in the right way, and still those beliefs can be true.

Now here's the central point: It is possible to argue that (Ignorance Inflation) is true even when *The Matrix* scenario is understood in the revised way in which we are now understandfing it. To see how this argument would go, however, we need to think a little about *higher-order knowledge*.

Higher order knowledge is knowledge about what we know. For example, If I know that I know that $2 + 2 = 4$ than I have one-bit of higher order knowledge. If I also know that I know that my wife loves me that I have another bit of higher order knowledge. Now, a key question about higher-order knowledge is whether we we *always* have it. In other words, the question is whether the following claim is true:

(KK) If *S* knows that *p*, then *S* knows that she knows that *p*.

Some philosophers think that (KK), or something very similar to it, just has to be true. They often say that if (KK) were false, we would be equally likely to draw infrences from things we do not know as from things we do know. Other philosophers think that (KK) just has to be false. These philosophers often say that (KK) has counterexamples. One type of these proposed counterexample involves chicken sexers - people who can reliably determine if a chick is male or female, but cannot explain how they do it. Philosophers suggest that the fact that chicken-sexers are reliabile at telling the sex of chickens allows them to know (e.g.) that this or that chick is male, but without knowing that they know that it is male. They don't know that they know it is male because they cannot rule out that they are just lucky guessers of the chick's sex.

Whatever you make of (KK), it can be used to support (Ignorance Inflation) by means of the following argument:

(Hypothesis) YOU KNOW THAT YOU ARE A MEMBER OF BEN GURION UNIVERSITY.

(Since KK is true) You know that YOU KNOW THAT YOU ARE A MEMBER OF BEN GURION UNIVERSITY.

(Since The Matrix is Bad) That YOU KNOW THAT YOU ARE A MEMBER OF BEN GURION UNIVERSITY entails that ***The Matrix* scenario is *not* real**.

(Since Knowledge Is Closed Under Entailment) If (i) that YOU KNOW THAT YOU ARE A MEMBER OF BEN GURION UNIVERSITY entails that ***The Matrix* scenario is *not* real** and (ii) you know that YOU KNOW THAT YOU ARE A MEMBER OF BEN GURION UNIVERSITY, then you know that ***The Matrix* scenario is *not* real**.

(So Hypothesis entails) You know that ***The Matrix* scenario is *not* real**.

(Excluded Matrix) Therefore, if YOU KNOW THAT YOU ARE A MEMBER OF BEN GURION UNIVERSITY, then you know that ***The Matrix* scenario is not real**.

(Ignorance Inflation*) Therefore, if you do *not* know that *The Matrix* scenario is not real, then you do *not* know that you are a member of Ben Gurion University.

Now notice that could run this argument not just with the claim that you are a member of Ben Gurion University, but also with the calim that you have a name that was given to you when you were young, or with the claim that you sometimes meet friends and family in person, etc. In fact, we could run this argument with any of the great many claims that you believe and which you would not know if *The Matrix* scenario were real. For this reason, we can deduce from the last argument that

(Ignorance Inflation) If you do not know that *The Matrix* scenario is not real, then you do not know a great many of the things you believe.

17.3 Relevant Alternatives

You might not know exactly what to think of these *arguments* for (Ignorance Inflation). But you might still suspect that (Ignorance Inflation) *itself* is implausible. Suppose you were an eye witness to a murder, and were asked to name the murderer in court. We would not usually accept the claim that since you do not know that The Matrix scenario is not real, you do not know who the murderer is. So does that suggest that (Ignorance Inflation) is false?

To answer this question, let's say that a *skeptical hypothesis* is a claim such that, if it is true, you do not know any (or at least, many) of the things you believe on the basis of your experience. For example, the claim that The Matrix scenario is real is a skeptical hypothesis. Now compare two views:

- Stroud's view: In order to know any (or at least, many) of the things you believe on the basis of your experience, you have to know that all skeptical hypotheses are not true.
- Relevant Alternatives Theory (Austin's view): *If* you have some *special reason* to believe that a skeptical hypothesis is true, then that skeptical hypothesis is *relevant*, and in order to know any (or at least, many) of the things you believe on the basis of your experience, you have to know that it is not true. More generally: Say that the claim that *a* is an *alternative to your knowing that p* if and only if the claim that *a* is such that, if *a* is the case, then you do not know that *p*. According to the relevant alternatives theory, if you have some *special reason* to believe that *a* is the case, then "*a*" is a *relevant alternative to your knowing that p*, and in order to know that *p*, you have to know that *a* is not the case.

(Ignorance Inflation) is a specific version of Stroud's view. On the other hand, there is no obvious way to get from the Relevant Alternatives Theory to (Ignorance Inflation). Therefore, the relevant alternatives theory points to a way of avoiding Cartesian skepticism. A difficulty for the Relevant Alternatives Theory, however, is the following line of thought: Regardless of whether it is relevant, if you do not know that some alternative to your knowing that *p* is not the case, then for all you know, some alternative to your knowing that *p is* the case. But if the alternative is the case, then for all you know, you do not know that *p* is the case. Therefore, by (KK) principle, you do not know that *p*.

Part VI
What Is Knowledge?

Texts:
- Gettier, E. (1963). "Is Justified True Belief Knowledge?". Analysis, 23, pp. 121-23.
- Zagzebski, L. (1994), "The Inescapability of Gettier Problems". The Philosophical Quarterly, 44(174), pp. 65-73.

18 Kinds of Knowledge

There are a few kinds of knowledge. In particular:

1. Knowledge of objects, or *acquaintance*. For example, knowing Lady Gaga, knowing Beethoven's 5th symphony, knowing the flavor of chocolate.
2. Knowledge of skill, or *knowing how*. For example, knowing how to walk, knowing how to ride a bike, knowing how to compose a symphony.
3. Knowledge of facts, or *knowledge that*, or *propositional knowledge*. For example, knowing that $1+1=2$, knowing that Beer-Sheva is south of Tel Aviv, knowing that I'm hungry.

In this class we will only be concerned with propositional knowledge.

19 The JTB View

The historical analysis of (propositional) knowledge says that

(JTB) *S* knows that *p* if and only if

(B) *S* believes that *p*

(T) it is true that *p*,

(J) S is justified in believing that *p*.

Let's go over each of these three conditions.

19.1 The Belief Condition, (B)

The belief condition says that you can only know what you believe. In other words, if you do not believe that *p*, you do not know that *p* either.

Note that epistemologists often distinguish between two kinds of belief:

- Outright belief. Outright belief is binary - you either believe that *p*, or you don't. Outright belief does not come in degrees. To outright believe that *p*, you have to be committed to its being the case that *p*.
- Credence, or degree of confidence. Credence is graded - your credence that *p* can be higher or lower, stronger or weaker.

The relationship between outright belief and credence is complicated, and we won't go into it here. All that matters for us is that when *the Belief Condition in the (JTB) view is a condition about outright belief*. This entails that you can only know what you outright believe, i.e., what you are committed to. You cannot know things you merely have some weak credence in.

19.2 The Truth Condition, (T)

Before we discuss the truth condition, we should discuss *factive attitude* in general. A factive attitude is an attitude that you can only bear to true propositions. In other words, if a proposition is false, you cannot bear any factive attitude to it. For example:

Factive Attitudes
- discovering that *p*
- being aware that *p*
- remembering that *p*
- perceiving that *p*
- *knowing that p*

Non-Factive Attitudes
- predicting that *p*
- being sure that *p*
- hoping that *p*
- seeming that *p*
- *believing that p*

The truth condition says that knowledge is a factive attitude: it says that you can only know truths, and what is false cannot be known. This idea can be confusing, so be aware that:

- Something can be true even if we do not know it, and even if we cannot know it. We cannot know whether there is a mountain of gold somewhere in the universe (this is because there are parts of the universe that we can never reach). But it can nevertheless be true that there is a mountain of gold. Nevertheless, if we *do* know something, that what we know is true.
- That knowledge is factive does not entail that if you know something, then you can prove that it is true or that your evidence guarantees that it is true. Many epistemologists believe that I can know that my car is out of gas by looking at the gas meter, even though, what I see in the meter does not allow me to prove, nor does it guarantee, that my car is out of gas.
- That knowledge is factive does not entail that if you believe you know something, then what you believe you know is true. For you may be wrong both about what is true, and about what you know.

19.3 The Justification Condition (J)

You can believe certain things, and your beliefs can even be true, and still you may not know those things. Examples:

- Suppose you flip a coin. Before it lands, you believe that it will land heads. You have no reason to believe that it will land heads. Your belief that it will land heads is irrational. Still, your belief is true - the coin will land heads. In this case, before the coin lands you have a true belief that is not knowledge.
- Consider the following passage from Plato's *Theaetetus* (200d-201d):

> *Socrates*: ...what are we going to say knowledge is?...
> Theaetetus: ...knowledge is true judgment. Judging truly is at least something free of mistakes, I take it, and everything that results from it is admirable and good. ...
> *Socrates*: Well, this won't take long to consider, anyway; there is a whole art indicating to you that knowledge is not what you say.
> Theaetetus: How's that? What art do you mean?
> *Socrates*: The art of the greatest representatives of wisdom—the men called orators and lawyers. These men, I take it, use their art to produce conviction not by teaching people, but by making them judge whatever they themselves choose. Or do you think there are any teachers so clever that within the short time allowed by the clock they can teach adequately to people who were not eye-witnesses the truth of what happened to people who have been robbed or assaulted?
> Theaetetus: No, I don't think they possibly could; but they might be able to persuade them.
> *Socrates*: And by "persuading them", you mean "causing them to judge", don't you?
> Theaetetus: Of course.
> *Socrates*: Then suppose a jury has been justly persuaded of some matter which only an eye-witness could know, and which cannot otherwise be known; suppose they come to their

> decision upon hearsay, forming a true judgment: then they have decided the case without knowledge, but, granted they did their job well, being correctly persuaded?
> THEAETETUS: Yes, certainly.
> *Socrates*: But, my dear lad, they couldn't have done that if true judgment is the same thing as knowledge; in that case the best juryman in the world couldn't form a correct judgment without knowledge. So it seems they must be different things.
> THEAETETUS: Oh, yes, Socrates, that's just what I once heard a man say; I had forgotten, but now it's coming back to me. He said that it is true judgment with an account that is knowledge; true judgment without an account falls outside of knowledge. ...

To solve this problem, we add the justification condition to the analysis of knowledge. The justification condition says that you can only know what you are justified in believing. In other words, if you are not justified in believing that *p*, you do not know that *p* either.

Note:

- *Epistemic Justification*. The justification that is mentioned in the justification condition is *epistemic justification*. To be epistemically justified in believing something is for your belief to be *epistemically appropriate*; it is for your belief to be proper given the concerns of epistemology. A more informative characterization of justification is difficult to give, both because the concept is supposed to be quite basic, and because the topic is very controversial. One main controversy is the one between internalists and externalists about epistemic justification:
  - *Internalism*. Internalism is the view that what a subject is epistemically justified to believe is determined entirely in virtue of the subject's internal (or, intrinsic) properties, i.e., entirely in virtue of the ways that the subject (and nothing else) is. On this view, necessarily, two subjects who are internally exactly similar are epistemically justified to believe the same things. In particular, two people who are internally exactly similar are epistemically justified to believe the same things, even if one of them is in the Matrix while the other is not.

    Two common forms of internalism are:
    - *Accessibilism*. Accessibilism is the view that what a subject is epistemically justified to believe is entirely determined by things to which the subject has some special sort of access (e.g., they are the subject's conscious states, or they are things the subject can know about to become aware of by introspection).
    - *Mentalism*. Mentalism is the view that what a subject is epistemically justified to believe is entirely determined by the subject's mental states.
  - *Externalism*. Externalism is the view that internalism is false. According to externalism, two people who are internally exactly similar, but who are different in that one of them is in the Matrix while the other is not, can be epistemically justified to believe different things.

  Despite the complications introduced by the internalism/externalism dispute, it is fair to say that the notion of an epistemically justified belief has something to do with notions like a reasonable belief, a belief held for good reasons, a belief held in accordance with the available evidence, a belief formed in a reliable way, etc.
- *Difference from Other Kinds of Justification*. Despite the complications introduced by the internalism/externalism dispute, we can better understand what epistemic justification is by distinguishing it from other kinds of justification, for example moral justification or pragmatic justification. For example, If I want to save a sick person in Larissa, which is to the north of me, and I believe that Larissa is to the north of me, then my belief is both morally and pragmatically justified - it will help me do something moral, which is to save a sick person in Larissa, and it will help me get what I want, which is to save a sick person in Larissa. But if I have no reasons to believe that Larissa is to the north of me, my belief may nevertheless by epistemically inappropriate and epistemically unjustified.
- *Justification and Truth*. Some epistemologists, but not all of them, hold that false beliefs can be epistemically justified. For example, some epistemologists hold that if the gas meter in my car says that it

is out of gas, and after looking at the meter I form the belief that my car is out of gas, then my belief is justified even if the gas meter is broken and my car does have gas in it.

- *Justification as a Kind Rather than an Activity*. Some epistemologists, but not all of them, hold that you can be epistemically justified in believing something, even if you cannot give a good argument for it. In other words, they hold that you can be epistemically justified in believing something, even if you cannot justify your belief. For example, some epistemologists hold that a young child who cannot yet explain herself in words can be justified in believing that Dad is standing in front of her, even if she cannot yet explain that she believes it for the reason that she sees her dad standing in front of her.

- *Doxastic vs. Propositional Justification*. Epistemologist Jim Pryor considers a case where he goes up to his ex-wife and tells her he can't live without her, she should leave her husband Tom and come back to him. She laughs in his face and tells him she doesn't care about him anymore, she's totally in love with her husband Tom. Pryor notes that at this point he has very good reasons for believing that his ex-wife does not love him. Even so, he might still refuse to believe it. He might think (unreasonably), "She still loves me, she just wants me to be jealous." In this case, Pryor has *propositional justification for believing* that his ex-wife does not love him.

  – Propositional justification for believing something is justification for forming or maintaining a belief, whether or not one does indeed have or form that belief.

  Now Pryor asks us to suppose that he goes home and asks a Magic 8 Ball, "Does my ex-wife will love me?" After the Magic 8 Ball says "Definitely not", Pryor trusts it and so forms the belief that his ex-wife does not love him. Pryor's new belief is defective. What is defective about it is that Pryor has this belief that his ex-wife does not love him for the wrong reasons, rather than for the right ones. So Pryor's new belief is not epistemically appropriate, and not epistemically justified, even though (as we noted a second ago) Pryor has propositional justification for it. A different way of expressing this idea is by saying that Prior's belief is not *doxastically justified*, even though Pryor has propositional justification for it.

  – A doxastically justified belief is a belief that is formed or maintained in a justified way.

  When a belief is doxastically justified, the subject both has propositional justification for holding it, and holds the belief *on the basis of* that justification. The thing on the basis on which we believe something is called the *grounds* of that belief.

19.4 JTB

We have seen some strong arguments for thinking that

(weak JTB) you know that *p only if* you have an (epistemically and doxastically) justified true belief that *p*.

This has led many to accept the (JTB) view, which (as we saw) says that

(JTB) you know that *p if and only if* you have an (epistemically and doxastically) justified true belief that *p*.

However, in a famously short paper from 1963, Edmund Gettier argued that even if (weak JTB) is correct, (JTB) is not...

20 The Gettier Problem

A *Gettier case* is a case in which someone has an epistemically and doxastically justified belief that *p*, but intuitively does not know that *p*. Examples:

- *Havit/Nogot* (Lehrer). You know that your friend Nogot has owned a Ford for the past 20 years ago. Nogot has also just offered you a ride in a Ford, which you know was parked in front of his home for the last year. You therefore justifiably believe that one of your friends owns a Ford. Furthermore, your

justified belief that one of your friends owns a Ford is true. Havit is your friend, and although you have no clue of this fact, Havit does indeed own a Ford. So Havit makes your justified belief true. The only problem is that Nogot does *not* own a Ford, and you have no clue that Havit owns one. So intuitively, you do not know that one of your friends owns a Ford. You have a justified true belief, but you do not know.

- *Gindi Tower* (Feldman). Bob, whom you know to be generally honest, has told you he lives in Gindi Tower. He's always wearing a T-shirt that says he lives in Gindi Tower, he's always going on and on about how great living in Gindi Tower is, and so on. The only think is that you are not sure of what Bob is called. You only remember his name starts with a "B". So on the basis of this you come to justifiably believe that a person whose name starts with a "B" lives in Gindi Tower. Furthermore, your belief is true - Barbara lives in Gindi Tower, and her name does indeed start with a "B". The only problem is that you have no idea about Barbara, and that Bob is a faker, and doesn't really live in Gindi Tower. So intuitively, you do not know that a person whose name starts with a "B" lives in Gindi Tower. You have a justified true belief, but you do not know.
- *Barn Façade County* (Ginet, in Goldman). Henry is driving in the countryside. In full view for him is a barn, with all the characteristic features. Henry has excellent eyesight, sees the barn and the features, and forms the belief that it is a barn. Henry's belief is true and justified. However, unbeknownst to Henry, the county he has just entered is full of paper-maché facsimiles of barns. The facsimiles are not barns, but just façades that look like barns from the road. Henry has not encountered any facsimiles yet, and he is seeing the only real barn in the county. Still, had Henry seen a facsimile, he would have mistaken it for a barn. So intuitively, Henry does not know that a barn is in front of him. He thus has a justified true belief, but he does not know.

Note:

1. Many people think that these cases would not be counterexamples to the JTB view, if we assume that you can only justifiably believe something if your grounds guarantee its truth.

   Why? Well, because

   (a) In the Havit/Nogot case, your belief that one of your friends owns a Ford would not be justified, because your grounds do not guarantee the belief's truth.

   (b) In the Gindi Tower case, your belief that a person whose name starts with a "B" lives in Gindi Tower would not be justified, because your grounds do not guarantee the belief's truth.

   (c) In the Barn Façade County case, if you think that Henry's belief that the thing there is a barn is based on *seeing the barn*, then his belief is based on grounds that guarantee the belief's truth. But if his belief is based on *having an experience in which it appears to him that there is a barn there*, then his belief is based on grounds that guarantee the belief's truth; and so his belief would not be justified.

2. At any rate, there is a serious problem for the view that you can only justifiably believe something if your grounds guarantee its truth. The problem is this: The grounds we have for our beliefs (e.g., our scientific beliefs) often do not guarantee the truth of those beliefs. So all those belief would be unjustified...

3. Zagzebski claims that Gettier has uncovered a problem not just to the JTB view, but for *any* attempt to analyze knowledge as true belief that has some further property:

   > "As long as the property that putatively converts true belief into knowledge is analyzed in such a way that it is strongly linked with the truth, but does not guarantee it, it will always be possible to devise cases in which the link between such a property and the truth is broken but regained by accident"

   Zagzebski thinks that her central claim is true because she has a recipe for generating Gettier cases:

(a) Start with a case in which a false belief possesses the property you think converts true belief into knowledge. [Zagzebski says that such a case is possible, because we are assuming that the property does not guarantee the belief's truth.]

(b) Add an element of luck to the case, so that the belief now ends up being true, though it still possesses the property you think converts true belief into knowledge.

Part VII
Can Science Discover The Laws Of Nature? The Negative Answer

Texts:

- Hume, D. Section IV. In his Enquiry Concerning Human Understanding. From the Stephen Buckle edition, 2007, Cambridge: Cambridge University Press, pp. 28-40.
- Popper, C. (1985). The Problem of Induction. In Miller, E. (ed.), Popper Selections. Princeton, NJ: Princeton University Press, pp. 101-105. (Excerpt.)

21 Induction

Here are some examples of inductive arguments:

[Hebrew] 93 percent of Israelis speak Hebrew. Alice is an Israeli. Therefore Alice likely speaks Hebrew.

[Neptune] In the 1820s only the first 7 planets of the solar system were known. But the 7th planet, Uranus, was observed traveling along a different orbit than the one predicted by Newtonian mechanics along with the assumption that there are only 7 planets in the solar system. Astronomers Urbain Le Verrier and John Couch Adams independently suggested that the existence of an eighth planet would provide the best explanation of Uranus's orbit. Then, in 1846, astronomer Johann Galle (helped by Heinrich d'Arrest) found a the planet where where Le Verrier said it would be.

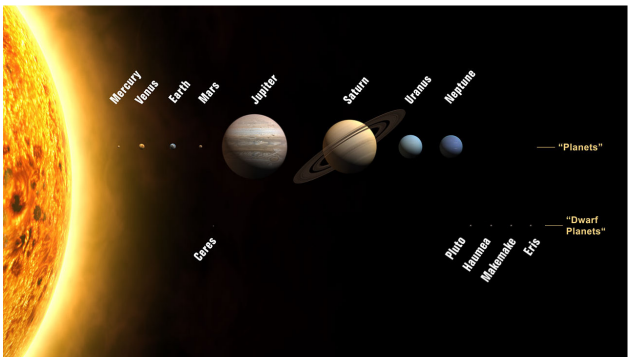

Figure 1: The Solar System. By WP - Planets2008.jpg, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=45708230

[Ravens] Every raven in a sample of 100,000 ravens is black. Therefore, all ravens are black.

[Aleh Yarok] 62 percent of voters in a random sample of 400 will vote for Aleh Yarok in the coming elections. So there is a 95% probability that between 58 percent and 66 percent of all voters will vote for Aleh Yarok in the coming elections.

[Bread Inference] All observed instances of bread have been nourishing. Therefore, all instances of bread will be nourishing.

All these arguments are inductive aruguments in the *broad sense* of induction:

(Broad sense) In the broad sense, an inductive argument is one in which the premises support the conclusion to some degree, but do not logically entail it.

However, only [Ravens], [Aleh Yarok], and [Bread Inference] are inductive arguments in the *narrow sense* of induction:

(Narrow sense) In the narrow sense, an inductive argument is an *enumeratively* inductive argument, i.e., an argument suggesting that since a sample has a property, the entire sampled population has that property too. [Or, more elaborately, that since r is the rate of A individuals in a random sample of size n, then there is degree of support d for holding that $r \pm q$ (where q is the margin of error) is the rate of A individuals in the sampled population.]

22 The Problem of Induction

We will focus on the narrow sense of induction, i.e., on enumerative induction. Hume thinks that all enumeratively inductive arguments assume something like the following Uniformity Principle:

[Uniformity Principle] Nature is uniform. Specifically, instances of which we have had no experience resemble those of which we have had experience.

The problem is that Hume thinks there is no argument for the [Uniformity Principle], and therefore, that no enumerative inductive argument should be accepted. A useful approximation of Hume's argument for this is the following:

P1 (Hume's fork). There are only two kinds of arguments: a priori arguments (= arguments whose conclusions are knowable independently of experience) and a posteriori arguments (= arguments whose conclusions are not knowable independently of experience).

Horn 1

P2. An a priori argument establishes a conclusion whose negation is impossible.

P3. The negation of the [Uniformity Principle] (= the claim that instances of which we have had no experience do *not* resemble those of which we have had experience) is possible.

C1. Therefore, there is no a priori argument for the [Uniformity Principle].

Horn 2

P4. Any a posteriori argument for the [Uniformity Principle] presupposes the [Uniformity Principle].

P5 (Non-circularity). An argument for a principle may not presuppose the same principle.

C2. Therefore, there is no a posteriori argument for the [Uniformity Principle].

C3. Therefore, there is no argument for the [Uniformity Principle].

P6. Enumerative inductive arguments presuppose the [Uniformity Principle].

C4. Therefore, no enumerative inductive argument (including, e.g., the [Bread Inference]) should be accepted.

Remarks:

- Those who accept C4 are *skeptics about enumerative induction*. Skepticism about enumerative induction is problematic since many laws of science (e.g., all organisms are made of cells; all emeralds are green, sugar dissolves in water...) are thought to be supported by enumerative inductive arguments. Therefore, skeptics about enumerative induction are at risk of being skeptics about laws of nature and thereby undermining the project of scientific exploration.
- There are other versions of Hume's argument too. Some of those avoid P6 and the [Uniformity Principle].
- Some ways out of the problem:
  - Deny P2, and say that contingent claims can be a priori. (Kant...)
  - Accept C1, but say that some a priori argument shows not that the [Uniformity Principle] is reasonable, or probable. (E.g., an a priori argument to the best explanation shows this.)

- Deny P4, and say that the [Uniformity Principle] can be non-circularly established by an a posteriori argument. On some views, e.g., the argument relies on a regress of inductive justifications, each using only empirical presuppositions.
- Deny P5, and say (e.g.) that a circular argument is acceptable in the case of justifying a fundamental form of reasoning.
- Deny C4 follows from C3 and P6 because we are rationally entitled to believe the [Uniformity Principle] due to role in reasoning.
- Accept C4 and say we can live with this.

23 Popper's Response

Popper is one of those who thought that one might accept C4 and live with it. How?

According to Popper:

- No enumerative inductive argument should be accepted, and nothing is justified by enumerative inductive arguments.
- Nevertheless, laws of nature can still be *rejected*, since rejecting a law only requires a *deductive* argument. The argument is that the law (along with starting and boundary conditions) deductively entails that a certain observation will be made. Therefore, if the observation is not made, the law (or the starting and boundary conditions) must be false.
- Laws of nature can also still be *accepted as tentative hypotheses*. This requires no argument.
- Since laws can still be accepted or rejected, the project of scientific exploration is not undermined.

Some concerns:

- The scientific project aims at knowledge of the laws of nature. Popper does not seem to allow for this.
- There are always many hypotheses which have not yet been refuted by the evidence, and these may contradict one another. Since none are yet falsified, Popper seems bound to say that they are all on an equal footing. Yet, scientists will typically want to say that one is better supported by the evidence than the others. Are the scientists being unreasonable?

Part VIII
Can Science Discover The Laws Of Nature? The Positive Answer

Texts:
- Mellor, D. H. (1991). "The Warrant of Induction". In his Matters of Metaphysics. Cambridge: Cambridge University Press, pp. 254-68.

24 Reminders

- To be epistemically justified in believing something is for your belief to be *epistemically appropriate*; it is for your belief to be proper given the concerns of epistemology.
- There is a controversy between internalists and externalists about epistemic justification:
  - *Internalism*. Internalism is the view that what a subject is epistemically justified to believe is determined entirely in virtue of the subject's internal (or, intrinsic) properties, i.e., entirely in virtue of the ways that the subject (and nothing else) is. On this view, necessarily, two subjects who are internally exactly similar are epistemically justified to believe the same things. In particular, two people who are internally exactly similar are epistemically justified to believe the same things, even if one of them is in the Matrix while the other is not.
  - *Externalism*. Externalism is the view that internalism is false. According to externalism, two people who are internally exactly similar, but who are different in that one of them is in the Matrix while the other is not, can be epistemically justified to believe different things.
- Today we shall look at an externalist answer to Hume's Problem of Induction. The externalist solution attempts to undermine the following assumptions in Hume's argument:

  P4. Any a posteriori argument for the [Uniformity Principle] presupposes the [Uniformity Principle],

  P6. Enumerative inductive arguments (including, e.g., the [Bread Inference]) presuppose the [Uniformity Principle],

  where the [Uniformity Principle] says that instances of which we have had no experience resemble those of which we have had experience; and where the [Bread Inference] is the argument that since all observed instances of bread have been nourishing, all instances of bread will be nourishing.

25 Mellor's Externalism About the Epistemic Justification of Beliefs

Mellor suggests that

- Epistemic justification is related to likelihood of truth.
- An observation makes a belief epistemically justified if the occurrence of the observation makes it likely that the belief is true.
- The occurrence of an observation will make it likely that a belief is true if facts that correspond to (and guarantee the truth of) the relevant kinds of beliefs are sufficiently likely to cause those beliefs.
- A belief of yours can be epistemically justified even if you do not know that it is epistemically justified.

Exercise 1 Apply this view to the case of the belief that there is a sparrow on the grass.

Exercise 2 Why is Mellor's view externalist?

26 Mellor's Externalism About the Epistemic Justification of Inferential Habits

Mellor suggests that

- An inferential habit is epistemically justified if it is likely for its conclusions to be true, given that the premises are true.
- It is likely for the conclusions of an inferential habit to be true (given that the premises are true) if a (deterministic or probabilistic) law of nature links facts that correspond to (and guarantee the truth of) the premises with facts that correspond to (and guarantee the truth of) the conclusion.
- An inferential habit of yours can be epistemically justified even if you do not know that it is epistemically justified.

Applying this to the case of the enumerative induction in the [Bread Inference], we get:

- The habit of inferring "since all observed instances of bread have been nourishing, all instances of bread will be nourishing" is epistemically justified if it is likely that all instances of bread are nourishing given that all observed instances of bread are nourishing.
- It is likely that all instances of bread are nourishing given that all observed instances of bread are nourishing if a (deterministic or probabilistic) law of nature links being bread with being nourishing.
- So, if a (deterministic or probabilistic) law of nature links being bread with being nourishing, then the habit of inferring "since all observed instances of bread have been nourishing, all instances of bread will be nourishing" is epistemically justified.

The central point here is this:

- According to Mellor, we may set out completely ignorant about whether a (deterministic or probabilistic) law of nature links being bread with being nourishing. Still, if such a law exists, its existence will make the habit of inferring "since all observed instances of bread have been nourishing, all instances of bread will be nourishing" epistemically justified. So, if it also happens to be the case that we are epistemically justified in believing that all observed instances of bread have been nourishing, we may well be epistemically justified in believing (the potential law of nature) that all instances of bread will be nourishing.

Exercise 3 What would Mellor say about P6? What would he say about P4?

Part IX
Do I Have Free Will? Am I Morally Responsible for My Actions? - Leeway Views

Texts:
- van Inwagen, P. (1998). The mystery of metaphysical freedom. In. van Inwagen, P. and Zimmerman, D. (eds.), Metaphysics: The Big Questions. Oxford: Blackwell, pp. 365-374.
- Frankfurt, H. (1969). "Alternate Possibilities and Moral Responsibility". Journal of Philosophy, 66, pp. 829-39.

27 Freedom and Moral Responsibility

27.1 Moral Responsibility

We can distinguish between two kinds of *actions* in this way:

- Actions for which the agent is morally responsible - actions that the agent performs in such a way, that she would deserve blame (praise) for performing them, if she understood that they were morally wrong (exemplary).
- Actions for which the agent is not morally responsible - actions that the agent performs in such a way, that she would not deserve blame (praise) for performing them, even if she understood that they were morally wrong (exemplary).

Example. Suppose you are at a party at a friend's house, when a muscle spasm causes you to spill wine over your friend's favorite couch. You are not morally responsible for spilling the wine, since you performed this action because of a spasm, which is not a way of performing an action that makes yo deserving of blame (even though you understood all along that it is morally wrong to spill wine on friends' couches.)

27.2 Freedom

We can also distinguish between two kinds of actions in this way:

- Free actions (or, actions that are performed freely) - actions that the agent *controls*; that are *up to* the agent.
- Non-free actions (or, actions that are not performed freely) - actions that the agent does *not* control; that are *not* up to the agent.

Example. Suppose you are at a party at a friend's house, when a muscle spasm causes you to spill wine over your friend's favorite couch. Then, the spilling of the wine is a non-free action, since it was not up to you that you spilled it; you did not control the spilling of the wine.

27.3 Debates

It is fairly widely *accepted* that freedom is necessary for moral responsibility, i.e., that

[Freedom necessary for responsibility] If some agent is morally responsible for an action, then the agent performed the action freely.

Example. Suppose you are at a party at a friend's house, when a muscle spasm causes you to spill wine over your friend's favorite couch. If freedom is necessary for moral responsibility, then since the spilling of the wine is a non-free action, you are not morally responsible for spilling the wine.

Apart from this widely held view, however, there are great *controversies* about the concept of freedom. Specifically, there are two main controversies about what conditions are necessary for freedom. The first controversy is over the truth of the claim that

[Alternatives necessary for freedom] If some the agent performed an action freely, then the agent was able to act otherwise (i.e., to perform an alternative action).

There are debates over (i) whether the claim is true, and (ii) whether the ability to do otherwise is compatible with determinism.

The second controversy is over the truth of the claim that

[Source-hood necessary for freedom] If some the agent performed an action freely, then the agent was the source of the action.

There are debates over (i) whether the claim is true, (ii) whether to be the source of an action is to be its ultimate originator, to be a merely appropriate cause of the action, or simply to be the performer of an action that does not have a sufficient cause (iii) what each of last three options requires, and (iv) whether source-hood is compatible with determinism..

28 Determinism

The two debates just mentioned about the concept of freedom are often discussed in connection with the topic of determinism. Determinism is the claim that the laws of nature and the past determine the future, or more precisely, it is the claim that

[Determinism] For any time instant *t*, it is metaphysically necessary that $((STATE_t \wedge LAWS) \supset FUTURE_t)$;

where "$STATE_t$" is a complete description of the state of the world *at t*, "*LAWS*" is the conjunction of all the laws of nature, and "$FUTURE_t$" is a complete description of the state of the world at any time *after t*.

Note:

1. If determinism is true, then there is exactly one future course of events that is compatible with the past and laws.
2. In the context of discussions of determinism, to say that an action or event *x* is *determined* is to say that it is metaphysically necessary that $((Y \wedge LAWS) \supset X)$; where "*Y*" is a complete description of something that occurred prior to *x*, "*LAWS*" is the conjunction of all the laws of nature, and "*X*" is a description of *x*.

Example. Suppose that my cat is chasing a butterfly now, and that the laws of nature dictate that if my cat is chasing a butterfly, that butterfly will necessarily get eaten. Also suppose that the present state of the world along with the laws of nature do *not* dictate whether the butterfly will try to run away from my cat. Then, since some future event is not determined by the present state of the world along with the laws of nature, determinism is false. Nevertheless, it is determined that the butterfly will get eaten.

29 The Problem of Freedom and Determinism, and Leeway-Incompatibilism

The central question, then, is how determinism relates to freedom and moral responsibility. There are lots of possible views here, but the ones relevant for our purposes are listed in the following chart:

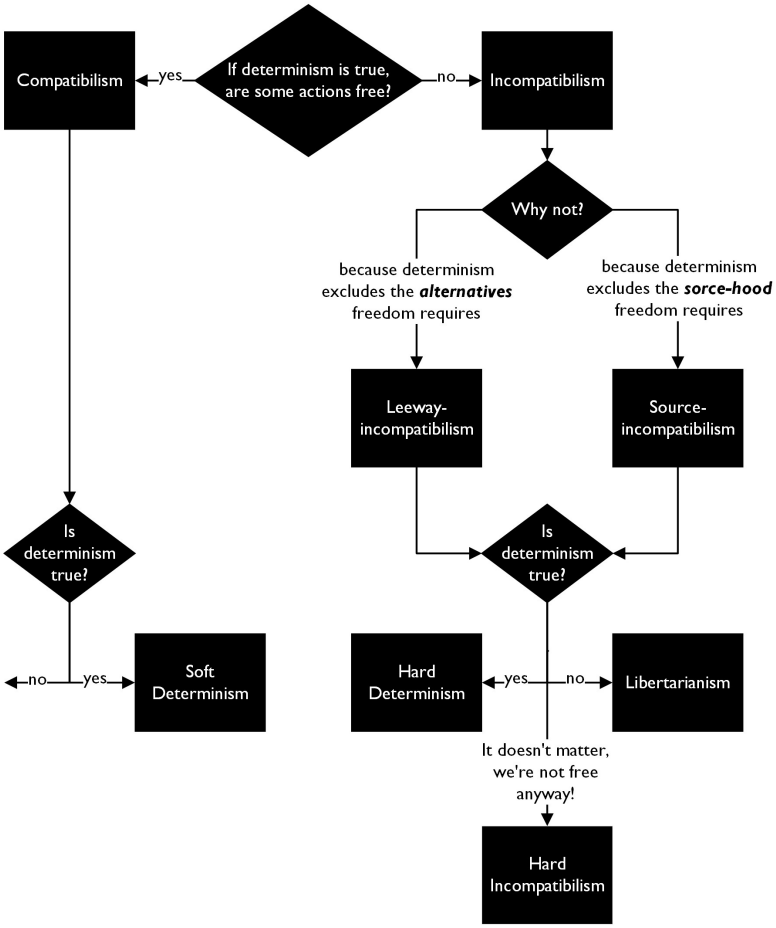

Figure 2: Views on the relation between determinism and freedom

We start with perhaps the two oldest arguments in this field of philosophy, which support the leeway-incompatibilist position. The first argument argues for incompatibilism on the grounds that freedom requires alternatives:

(Alternatives necessary for freedom) If some agent performed an action freely, then the agent was able to act otherwise (i.e., to perform an alternative action).

(Determinism excludes alternatives) If determinism is true, then no agent was ever able to act otherwise.

(Incompatibilism) Therefore, if determinism is true, no agent ever performed an action freely.

The second argument argues that determinism is incompatible with moral responsibility, again on the grounds that freedom requires alternatives:

(Freedom necessary for responsibility) If some agent is morally responsible for an action, then the agent performed the action freely.

(Alternatives necessary for freedom) If some the agent performed an action freely, then the agent was able to act otherwise (i.e., to perform an alternative action).

(Principle of alternate possibilities [= PAP]) If some agent is morally responsible for an action, then the agent was able to act otherwise (i.e., to perform an alternative action).

(Determinism excludes alternatives) If determinism is true, then no agent was ever able to act otherwise.

(Determinism excludes moral responsibility) Therefore, if determinism is true, no agent ever performed an action freely.

Every premise in these arguments is disputed. Today we shall focus on a particularly famous attack on PAP, due to Frankfurt.

30 Frankfurt's Argument Against Leeway-Incompatibilism and PAP

Frankfurt aims to refute [PAP] by counterexample. He describes the following case:

> Suppose someone—Black, let us say—wants Jones_4 to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones_4 is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones_4 is going to decide to do something other than what he wants him to do. If it does become clear that Jones_4 is going to decide to do something else, Black takes effective steps to ensure that Jones_4 decides to do, and that he does do, what he wants him to do. Whatever Jones_4's initial preferences and inclinations, then, Black will have his way.
>
> What steps will Black take, if he believes he must take steps, in order to ensure that Jones_4 decides and acts as he wishes? ... let Black manipulate the minute processes of Jones_4's brain and nervous system in some more direct way, so that causal forces running in and out of his synapses and along the poor man's nerves determine that he chooses to act and that he does act in the one way and not in any other. ...
>
> Now suppose that Black never has to show his hand because Jones_4, for reasons of his own, decides to perform and does perform the very action Black wants him to perform.

Now Frankfurt reasons as follows:

- Jones_4 is unable to decide and do otherwise than Black wants him to.
  Reason. Either Jones_4 is inclined to decide as Black wants him to, or he is not so inclined. If he is inclined to decide as Black wants him to, then he will decide and then do as as Black wants him to. If he is not so inclined, then Black would interfere and again Jones_4 will decide and then do as Black wants him to. So either way, Jones_4 will decide and then do as Black wants him to. So Jones_4 cannot avoid deciding and doing as Black wants him to.
- Jones_4 is morally responsible for deciding and doing as Black wants him to.
  Reason. Jones_4 decided and acted on his own, without anyone interfering with him.
- The principle of alternate possibilities is false, i.e., it is false that if some agent is morally responsible for an action, then the agent was able to act otherwise.
  Reason. The Jones_4-case is a case in which an agent is (1) unable to act otherwise than they do, and (2) morally responsible for the action.

A possible worry for Frankfurt's argument is this: Jones might be able to decide otherwise than Black wants him to. This is because after Black decided *not* to interfere (becauseJones_4 was inclined to decide as Black wants), Jones_4 could still have changed his mind and decided something else. Concerns of this nature led to more developed examples and much further debate.

Part X
Do I Have Free Will? Am I Morally Responsible for My Actions? - Source, Libertarian and Hard Incompatibilist Views

Text:

- Pereboom, D. (2011). Free Will Skepticism and Meaning in Life. In Kane, R. (ed.), Oxford Handbook of Free Will, Second edition. New York: Oxford University Press, pp. 407–24.

31 The Problem of Freedom and Determinism, and Source-incompatibilism

Suppose you accept Frankfurt's argument, and so deny PAP. Must you then be a compatibilist? No. After all, it is still possible that freedom requires source-hood of a kind that is incompatible with determinism. We call this kind of source-hood "ultimate origination".

Pereboom is one of those who believe that freedom requires "ultimate origination". We'll start with his argument against compatibilism, the so-called "multiple-case manipulation argument". The argument involves the following four cases:

> *Case 1.* Neuroscientists can manipulate Professor Plum's mental state at any moment through the use of radio-like technology. At a certain time, when just before Prof. Plum begins to deliberate about whether to kill Ms. White for some personal advantage, they press a button that causes Plum's deliberative process to be egoistic. The neuroscientists know that this will deterministically result in Plum's deciding to kill White. Now, when Plum decides to kill her, he does not think and act contrary to character, because his reasoning processes are often rationally egoistic. His effective first-order desire to kill White conforms to his second-order desires. The process of deliberation from which his action results is reasons-responsive; in particular, this type of process would have resulted in his refraining from killing White in some situations in which the egoistic reasons were different. Still, he is not exclusively rationally egoistic, because he typically regulates his behavior by moral reasons when the egoistic reasons are relatively weak. He is also not constrained, in the sense that he does not act because of an irresistible desire. The neuroscientists do not induce a desire of this kind.
>
> *Case 2.* Like case 1, except that the neuroscientists don't have the radio-wave technology, and instead program Plum at the beginning of his life to weigh reasons for action so that he is often but not exclusively rationally egoistic, with the consequence that in the circumstances in which he now finds himself, he is causally determined to engage in the reasons-responsive process of deliberation and to have the set of first and second-order desires that result in his killing White.
>
> *Case 3.* Like case 2, except that instead of neuroscientists programming Plum, Plum undergoes rigorous training by his household and community so that he is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1 and 2). This training takes place when Plum is too young to have the ability to prevent or alter the practices that determine his character.
>
> Case 4. Like case 3, except that instead of rigorous training, it is the past history of the world and laws of nature that cause him to be often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1 and 2).

The argument is this: Intuitively, Plum is *not* morally responsible for deciding to kill Ms. White in case 1; and that since there is no difference relevant to Plum's moral responsibility between cases 1 and 2, 2 and 3, and 3 and 4, Plum is *not* morally responsible for deciding to kill Ms. White in case 4 either. But in case 4, Plum is no different from any agent who is determined to perform a selfish action. So if determinism is true (and all actions are determined), no agent is morally responsible for her selfish actions.

Responses to the "multiple-case manipulation argument come in two kinds: *Hard-line responses* say that Plum *is* morally responsible for deciding to kill Ms. White in case 1. *Soft-line responses* say that there is a relevant difference between a certain pair of neighboring cases. We'll look at McKenna's hard line response. According to McKenna, we should start by considering case 4. McKenna highlights that the agent in the case of ordinary determinism satisfies all compatibilist relevant conditions for moral responsibility and free action, and it is therefore not obvious that the agent is not a morally responsible agent. He then uses Pereboom's generalization strategy to proceed through the remaining cases. Given Pereboom's assumption that there are no compatibilist relevant differences between the cases the intuition that it is not obvious that the agent lacks moral responsibility transfers down the line until the case of moment-to-moment manipulation. McKenna then points out that, given the generalization strategy and the lack of compatibilist-responsibility relevant differences between the cases, even the agent manipulated from moment-to-moment could be a morally responsible, free-acting agent.

32 Libertarianism and the Problem of Luck

Suppose you accept Pereboom's multiple-case manipulation argument, and so believe that freedom is incompatible with determinism. What do you think about our actions - are some of them free or not? Libertarians say that (i) determinism is false, and (ii) that if determinism is false, some of our actions are free. This Libertarian view entails that some of our actions are free. But Pereboom rejects Libertarianism. He rejects it because he thinks that (ii) is false, i.e., that *even if* determinism is false, none of our actions are free. One way to present his argument for this view can be put as follows:

1. Leeway-Incompatibilism and compatibilism are false.
2. Therefore, source-insompatibilism is true.
3. Therefore, freedom requires ultimate-origination, i.e., if some agent performed an action freely, then the agent was the ultimate originator of the action (which entails that the action was not determined).
4. If an agent is the ultimate originator of some action, the agent is somehow involved in causing the action (but without the action being determined).
5. An agent can be involved in causing an action without the action being determined in just two mutually exclusive ways:
   (a) either the agent causes the action herself (i.e., not only by undergoing prior events),
   (b) or the action is caused—but is not determined—by prior events the agent undergoes; and the action is not caused by the agent herself.
6. The idea that agents cause actions by themselves is implausible.

   (No massive coincidences) Why accept claim 6? Well, our best physical theories suggest that actions are either determined by past events, or are caused by past events with a certain fixed probability. Therefore, if agents caused actions themselves, it would be a massive coincidence if those actions were performed exactly when, or with the exact probability, that our best physical theories predict. Since we should not expect such massive coincidences to occur, and since we should accept what our best physical theories predict, we should not hold that agents cause actions themselves, independently of past events.
7. Therefore, claim (5a) is false, and claim (5b) is true.
8. Actions that are caused—but are not determined—by prior events the agent undergoes (and not caused by agents themselves) are not free.

   (The Problem of Luck) Why accept claim 7? Well, suppose an action of Alice's is not determined. What settles that Alice performs the action, i.e., what ensures that Alice perform the action? Well, if the action is neither determined nor caused by Alice herself, nothing ensures that Alice performs

the action. In this sense, Alice disappears from performing the action. Therefore, Alice cannot reasonably be the ultimate originator of the action. Since Alice's action would be free only if Alice were it's ultimate originator, Alice's action is not free. And for similar reasons, all actions that are caused—but are not determined—by prior events the agent undergoes (and not caused by agents themselves) are not free.

9. Therefore, none of our actions are free.

Pereboom's conclusion suggests not only that libertarianism is false, but also that we are not free and so presumably also not morally responsible for our actions. This is the hard-incompatibilist position. But can we live with this position?

33 Living with Hard-incompatibilism

Pereboom thinks that we *can* live with hard-incompatibilism, but that this will require some changes to our normal practices. In particular, he thinks that

- We would be morally required to stop *praising* and *blaming* people for their actions.
- We would be morally required to stop *resenting* others, being *morally angry* with them, and it would be irrational to feel *guilt*.
- We would be morally required to stop *retaliating* against wrongdoers.
- We would be morally required to stop *deterring* people from doing wrong, by inflicting suffering on wrongdoers.

Nevertheless, he also thinks that

- We could still hold that some actions are *wrong*/bad, and that some actions are *right*/good.
- We could still *educate* people by telling them which actions are wrong and which are right, and by encouraging them to perform the right ones and avoid the wrong ones.
- We could still *prevent* dangerous criminals from posing a danger to themselves or to society, but we would have to do so by means that are *proportional* to the danger people pose, that inflict *minimal suffering*, and that facilitate their becoming less dangerous (i.e., that are *rehabilitative*).
- We could still recognize people's *achievements*, in the sense that we can still recognize that people have accomplished their goals through effort. (However, they people not deserve praise for this.)
- We could still retain the sense of our *self-worth*, as this worth is not due to our praiseworthiness to a large extent.
- We could still feel *sad* or alarmed about events, *resolve* to do right and not wrong, *forgive* wrongdoings (in the sense of stopping to take them as reasons to undermine relationships), and express *gratitude* (in the sense of expressing thanks and being happy for certain actions).
- We could still *love* others.

Part XI
Appendices

34 Glossary

counterfactual conditional A counterfactual. *see* counterfactual

haecceity An individual essence. *see* individual essence

35 English-Hebrew Dictionary

abstract מופשט
abundant property תכונה במובן השופע
accidental property תכונה מקרית
agent פועל [ולא סוכן! סוכן הוא מי שמשמש כשליחו של אחר. פועל הוא מי שמבצע פעולות.]
brute fact עובדה חסרת הסבר
bundle צרור, צביר
categorical property תכונה קטגורית, תכונה בלתי-מותנית
cause גורם, סיבה
characterization principle עקרון האפיון
class אוסף, קבוצה
composite שלם מורכב
concrete קונקרטי, גשמי
constituent חלק משלם
constituent ontology אונתולוגיה של שלמים וחלקים
contiguity סמיכות בחלל-זמן
cosmological argument טיעון קוסמולוגי
counterfactual פסוק תנאי נוגד מציאות [פסוק מהצורה "לו היה___ אז היה ___", כאשר אחרי "לו" מופיע תנאי שאינו מתקיים]
determination קביעה, היקבעות
determinism דטרמיניזם, היקבעות על ידי אירועי העבר הרחוק וחוקי הטבע
disposition דיספוזיציה, כח
effect אפקט, תולדה, תוצאה
essential property תכונה מהותית
event אירוע
event-determination קביעה על ידי אירועים
existent דבר קיים
extensionality principle עקרון האקסטנסיונליות
extrinsic property תכונה אקסטרינזית, תכונה חיצונית
fine-tuning כוונון עדין
freedom to do otherwise יכולת לעשות אחרת
fundamental יסודי
haecceity הקסאיטי, מהות אינדיבידואלית
immanent אימננטי, חלל-זמני
indeterminism אינדטרמיניזם, אי-היקבעות על ידי אירוע העבר הרחוק וחוקי הטבע
individual אינדיבידואל, עצם
individual essence מהות אינדיבידואלית, הקסאיטי
intrinsic property תכונה אינטרינזית, תכונה פנימית
irreducible לא בר-רדוקציה
mereology תורת היחס חלק מ-
monadic trope טרופ מונאדי, טרופ חד מקומי
monadic universal אוניברסל מונאדי, אוניברסל חד-מקומי
monocategorical ontology אונתולוגיה חד-קטגורית
natural טבעי
nature טבע
nominalism נומינליזם
nomological של, או נוגע ל-, חוקים
non-instantiable לא בר-מימוש
object אובייקט [מונח רב משמעי]
see: existent, individual, non-instantiable, particular
ontological argument טיעון אונתולוגי
ontological category קטגוריה אונתולוגית
ontology אונתולוגיה, תיאוריה אונתולוגית
over-determination היקבעות-יתר
particular פרטיקולר
polyadic trope טרופ פוליאדי, טרופ רב-מקומי
polyadic universal אוניברסל פוליאדי, אוניברסל רב-מקומי
polycategorical ontology אונתולוגיה רב קטגורית
preemption מניעה מראש
primary ontological category קטגוריה אונתולוגית ראשית
primitive לא בר-ניתוח
property תכונה, איכות, מאפיין
reasons-responsiveness רגישות לטעמים, רגישות לנימוקים
reduction רדוקציה, העמדה על, זיהוי עם משהו מאוסף רחב יותר, גזירה מתורה כללית יותר
regularity רגולריות, סדירות
rejoinder תשובה לתשובה
relation יחס
relational ontology אונתולוגיה של יחסים
sequence סדרה
self-determination קביעה עצמית
sparse property תכונה במובן הדל
spurious מזויף
substance סובסטנציה [המונח לרוב מתורגם "עצם", אך תרגום זה שמור אצלנו כחלופה למילה "אינדיבידואל"]
see: individual
substrate סובסטרט, מצע
subsumption הכלה תחת
surrejoinder תשובה לתשובה
transcendent טרנסצנדנטי, לא-חלל-זמני
trumping הבסה
tu quoque גם את.ה [לטינית]
ultimate origination היות מקור אחרון

36 Hebrew-English Dictionary

אובייקט object (could mean either existent, individual, non-instantiable, or particular)
אונתולוגיה של שלמים וחלקים constituent ontology
אוניברסל חד-מקומי monadic universal
אוניברסל מונאדי monadic universal
אוניברסל פוליאדי polyadic universal
אוניברסל רב-מקומי polyadic universal
אונתולוגיה ontology
אונתולוגיה חד-קטגורית monocategorical ontology
אונתולוגיה רב קטגורית polycategorical ontology
אונתולוגיה של יחסים relational ontology
אוסף class
איכות property
אימננטי immanent
אינדטרמיניזם indeterminism
אינדיבידואל individual
אירוע event
אפקט effect
גורם cause
גם את.ה tu quoque [Latin]
גשמי concrete
דבר קיים existent
דטרמיניזם determinism
דיספוזיציה disposition
הבסה trumping
היות מקור אחרון ultimate origination
היקבעות determination
היקבעות-יתר over-determination
הכלה תחת subsumption
העמדה על reduction
הקסאיטי haecceity, individual essence
חוקים; של, או נוגע ל- nomological
חלל-זמני immanent
חלק משלם constituent
טבע nature
טבעי natural
טיעון אונתולוגי ontological argument
טיעון קוסמולוגי cosmological argument
טרופ חד מקומי monadic trope
טרופ מונאדי monadic trope
טרופ פוליאדי polyadic trope
טרופ רב-מקומי polyadic trope
טרנסצנדנטי transcendent
יחס relation
יכולת לעשות אחרת freedom to do otherwise
יסודי fundamental
כוונון עדין fine-tuning
כח disposition
לא בר-מימוש non-instantiable
לא בר-ניתוח primitive
לא בר-רדוקציה irreducible
לא-חלל-זמני transcendent

מאפיין property
מהות אינדיבידואלית haecceity, individual essence
מופשט abstract
מזויף spurious
מניעה מראש preemption
מצע substrate
מראולוגיה mereology
נומינליזם nominalism
סובסטנציה substance
סובסטרט substrate
סדירות regularity
סדרה sequence
סיבה cause
סמיכות בחלל-זמן contiguity
עובדה חסרת הסבר brute fact
עקרון האפיון characterization principle
עקרון האקסטנסיונליות extensionality principle
עצם individual
פועל agent
פסוק תנאי נוגד מציאות counterfactual, counterfactual conditional
פרטיקולר particular
צביר bundle
צרור bundle
קבוצה class [although in most contexts it is better to translate "set"]
קביעה determination
קביעה על ידי אירועים event-determination
קביעה עצמית self-determination
קונקרטי concrete
קטגוריה אונתולוגית ontological category
קטגוריה אונתולוגית ראשית primary ontological category
רגולריות regularity
רגישות לטעמים reasons-responsiveness
רגישות לנימוקים reasons-responsiveness
רדוקציה reduction
שלם מורכב composite
תולדה effect
תוצאה effect
תורת היחס חלק מ- mereology
תיאוריה אונתולוגית ontology
תכונה property
תכונה אינטרינזית intrinsic property
תכונה אקסטרינזית extrinsic property
תכונה בלתי-מותנית categorical property
תכונה במובן הדל sparse property
תכונה במובן השופע abundant property
תכונה חיצונית extrinsic property
תכונה מהותית essential property
תכונה מקרית accidental property
תכונה פנימית intrinsic property

תכונה קטגורית categorical property
תשובה לתשובה rejoinder

תשובה לתשובה לתשובה surrejoinder