Phil. 2750
Notes #1: Course Requirements, Introduction

To discuss today:
This course: requirements, subject, guidelines
What is philosophy?
What does science fiction have to do with philosophy?
What should you learn from this class?

I. About this Class

Review syllabus. Some highlights:
- Who should take this class?
  - Class will contain:
    Lots of arguments
    Theoretical, abstract questions
    Controversial ideas
  - Will not contain:
    Directly practical knowledge
    Touchy-feely, self-esteem-boosting material
- Course requirements. Quizzes, paper, participation
- Miscellaneous guidelines:
  Come on time.
  Come to office hours.
  Participate.
- What should you do now?
  Get books & reserve readings <http://libraries.colorado.edu/search/p?SEARCH=huemer>.
  Read the syllabus.
  Read the Pollock selection.

II. What Is Philosophy?

A. The Subject Matter of Philosophy
- Philosophy studies some general, fundamental questions, about the nature of the world and our place in it.
- Three main branches:
  1. Metaphysics - Studies what sorts of things in general exist, and what sort of world this is. Examples: existence of God, free will vs. determinism, distinction between body and soul, the Ship of Theseus question.
  2. Epistemology - Studies the nature of knowledge - what is it and how do we know what we know?
  3. Ethics - Studies evaluative questions - what is good, bad, right, and wrong.
- Some smaller branches of philosophy:

4. Political philosophy - Studies the source of political authority, the best overall structure for society and the state, and related questions. (Can be seen as a branch of ethics.)
5. Aesthetics - Studies the nature of art, beauty, and related questions. (More generally: the nature of aesthetic qualities.)
6. Logic - Studies reasoning, esp. the principles of correct reasoning. Closely related to, but not the same as, epistemology.

B. The Methods of Philosophy

Philosophy in the Western tradition mainly relies on logical arguments & common experience.

C. An Example of Philosophy

The Ship of Theseus:

Theseus sailed around the Mediterranean for 10 years. During this time, he periodically had to replace pieces of the ship due to wear and tear. After 10 years, every plank in the ship had been replaced, one at a time. Q: Was it still the same ship?

Notice things about this question:

- A priori, not empirical.
- Far-reaching implications (identity of composite objects over time).
- Puzzling. Compelling arguments for incompatible positions.

III. Science Fiction & Philosophy

• Science fiction is
  a) Imaginative, yet
  b) Realistic. (See how both of these are the case.)
• Provides material for thought experiments.
  - Exploring concepts through interesting cases on the borderline. Ex.:
    The concept of persons and: robots, alien species, symbiotic species, person-splitting, machine/organic hybrids.
    See Asimov's robots, Data, the Trill (from ST), "Second Chances" (ST:TNG), Borg
  - Exploring human nature through giving people special abilities. Ex.:
    Ability to rewrite the past, immortality, godlike powers over others.
    See ST:TOS, "Where No Man Has Gone Before"; Q from ST:TNG; various time travel stories.
• Some SF possibilities may actually be realized.
  - Androids, human/machine hybrids (like the Borg), nanotechnology (as in The Diamond Age), sophisticated virtual reality (Snow Crash), genetic engineering (Gattaca), colonies on the moon (The Moon Is a Harsh Mistress), radical restructuring of society (Diamond Age; Brave New World).
  - Only nerds will be ready.

IV. What I Hope You Learn from this Class

A. Some Important Philosophical Ideas
  - About free will, consciousness, personhood, time, knowledge, human nature, and the structure

of society.

B. Thinking Skills

- Philosophy teaches us to think more clearly, to avoid common confusions.
- Philosophy teaches us to reason more cogently, to avoid common fallacies.
- Philosophy makes us aware of the fundamental questions.

C. Philosophical Attitude

The Cardinal Rule of Philosophy: Truth comes first.

When doing philosophy, we are trying to identify what is true. That comes before personalities, feelings, and desires. The following rules are all consequences of this.

Four Principles of Philosophical Comportment:

1. Philosophers question:
   - Question the claims of others.
   - Question your own beliefs.
   - This does not mean refusing to accept anything as true!
2. Philosophy is impersonal:
   - The philosopher does not choose beliefs based on his personality or feelings.
   - The philosopher does not take criticism of ideas personally.
   - The philosopher does not accept or reject philosophical claims based on who says them.
   - The philosopher does not go along with ideas because of personal or social consequences of criticizing them.
3. Philosophers are guided by reason:
   - The philosopher has reasons for his beliefs.
   - The philosopher asks for the reasons for others' beliefs.
   - The philosopher is moved by good reasons presented to him.
4. Philosophers are open-minded and critical:
   - Our ideas and arguments are open to criticism. The philosopher looks for objections to his beliefs.
   - The ideas and arguments of others are also open to criticism.

Phil. 2750
Notes #2: Cartesian Skepticism

I. What Is Cartesian Skepticism?

Skepticism: Roughly, any philosophical view according to which some large class of things we normally believe (a) we do not know, or (b) we are not justified in believing.

Varieties of skepticism:

- External World Skepticism: We cannot have knowledge/justified belief about any contingent truths about the external world.
  The external world: That which is independent of one's own mind.
  Contingent truths: Things that could (conceivably) have been otherwise.
- Global Skepticism: We cannot have knowledge/justified belief about anything at all.
- In this unit, we consider external world skepticism regarding justified belief. This is the view that we have no justification for (contingent) claims about the external world.

Cartesian Skepticism:

- This is named after René Descartes. (See: Meditations on First Philosophy.)
- "Cartesian" skeptical arguments involve "skeptical scenarios":
  a) Scenario in which everything appears as it actually does, but
  b) Your beliefs are radically mistaken.
- Examples:
  - The dream scenario
  - The deceiving God
  - The brain in a vat
  - The Matrix, etc.

II. The Brain in a Vat argument

1. If you know that P and P entails Q, then you can know that Q. (Premise: Closure Principle for knowledge.)
2. You can't know you're not a BIV. Argument for this:
   a. Our sensory experiences are the only evidence we can have for claims about the external world. (Premise.)
   b. If you were a BIV, you would have the same sort of sensory experiences as you actually have. (Premise.)
   c. Your experiences are not evidence that you're not a BIV. (From b.)
   d. You cannot have evidence that you're not a BIV. (From a, c.)
   e. You can't know you're not a BIV. (From d.)
3. Therefore, you don't know (for example) that you have two hands. (From 1, 2.)
   (Implicit: Having 2 hands entails not being a BIV.)

In the following classes, think about which premise (if any) is denied by each response to skepticism.

III. Moorean Responses to Skepticism

Moore's Proof of an External World

We all know there are external objects. But is it possible to prove this? Moore says it's easy:

1. Here is one hand. (Gesture.)
2. Here is another hand. (Gesture.)
3. Therefore, there are external objects.

Moore & Pollock's general response to skeptical arguments

Skeptic's argument:

1. If you know that P and P entails Q, then you can know that Q.
2. You can't know you're not a BIV.
3. Therefore, you don't know (for example) that you have two hands.

Moorean argument:

1. If you know that P and P entails Q, then you can know that Q.
2. I know I have two hands.
3. So, I can know I'm not a BIV.

Schematically:

Skeptic:	Moore:
P_1	P_1
P_2	K
∴ ¬K	∴ ¬P_2

Which is better?

- Both equally valid.
- Both equally circular or non-circular.
- Which has the more plausible premise?
- Neutral presentation:
  - Three propositions are jointly incompatible: P_1, P_2, and K. At least one must be rejected.
  - Each initially seems true.
  - Solution: reject the one we are least confident of.
- Moore & Pollock's point generalizes to all skeptical arguments. Lesson:
  In a persuasive argument, each premise must be more initially plausible than the negation of the conclusion.

Phil. 2750
Notes #3: Are You Living in a Computer Simulation?

I. The Simulation Argument

An argument that we are probably living in a computer simulation

1. Functionalist theory of the mind. A sufficiently detailed simulation would be conscious.
2. An intelligent species has a good chance of reaching a "posthuman" stage.
3. There is a good chance that posthuman intelligent beings would run computer simulations.
4. If they did, there would be many more simulated beings than "real" beings.
5. The probability of being a simulated being = the fraction of all beings who are simulated beings. (The Pr. of being a real person = the fraction of all beings who are real people.)
6. Thus, there is a good chance that you are a simulated being.

Bostrom says: One of the following should be accepted:

a) that our species will probably go extinct before it reaches a much more advanced state;
b) that almost none of the highly advanced beings who could do so would choose to run sophisticated computer simulations; or:
c) that we're probably computer-simulated beings.

- This is the official thesis of the paper. He's not asserting (c).
- He is definitely asserting premises (1), (4), and (5).

II. Which of the Three Possibilities Is Most Likely?

Why think (a)?

- Posthuman stage: A state in which people are extremely technologically advanced. So-called because these beings may be beyond human, e.g., genetically engineered, human/machine hybrids, etc.
- If technological progress continues, this stage will occur, perhaps in a few decades or a few centuries.
- If intelligent species in general are unlikely to reach a posthuman stage, then our species will probably not reach it.
- Why might we fail to become posthuman?
  - Species extinction—war, diseases, other dangerous technology.
  - Collapse of technological civilization.

Why think (b)?

- Maybe simulations would be considered unethical.
  - Too much suffering?
  - Is it wrong to create us??
- Maybe posthumans would have entirely different interests.

Why think (c)?

- Posthumans would have vast computing power.

- "Moore's Law": computing power doubles about every 2 years.
- This has some theoretical limits. But:
  - ☞ A planet-sized computer could do 10^{42} operations per second.
  - ☞ Cost of a realistic simulation of human history: 10^{33}-10^{36} operations.
  - ☞ The planet-sized computer "could simulate the entire mental history of humankind . . . by using less than one millionth of its processing power for one second."

• Thus, the vast majority of beings with minds like us could be simulated beings.

III. Bostrom's Premises

Bostrom is asserting these:

1. Functionalist theory of the mind. A sufficiently detailed simulation would be conscious.
4. If posthumans ran ancestor simulations, then there would be many more simulated beings than "real" beings.
5. The probability of being a simulated being = the fraction of all beings (with experiences like ours) who are simulated.

Why accept (5)?

• Analogy: Suppose 2/3 of the population has genetic condition S. Given no other information, what is the probability that you have S?

• Suppose 1/4 of all babies result from unintended pregnancies. Given just this information, what is the probability that you resulted from an unintended pregnancy?

• What if all beings placed bets on whether they were simulated or not? What would be the fair betting odds?

• If 100% of all beings are simulated, what is the probability that you are simulated?

Why accept (4)?

• Discussed above.

Why accept (1)?

• "Arguments for this thesis have been given in the literature, and although it is not entirely uncontroversial, we shall here take it as a given." (2)

• "This attenuated version of substrate independence is quite widely accepted." (2)

Phil. 2750
Notes #4: The Matrix as Metaphysics

I. Thesis: The Matrix/Simulation Scenario Is Not a Skeptical Scenario.

1. The Matrix Hypothesis "is equivalent to" the Metaphysical Hypothesis.
2. The Metaphysical Hypothesis is not skeptical.
3. So the Matrix Hypothesis is not skeptical.

Matrix: A computer simulation of a world.
The Matrix Hypothesis: We are living in a matrix.

The Metaphysical Hypothesis:

a) The Creation Hypothesis: Physical spacetime and its contents were created by a being or beings outside spacetime.

b) The Computational Hypothesis: Microphysical processes are constituted by computational processes.

c) The Mind-Body Hypothesis: Our minds are outside physical spacetime but interact with processes in physical spacetime.

d) The computational processes were designed by the creators as a computer simulation of a world.

II. The Metaphysical Hypothesis Is Not Skeptical

a) The Creation Hypothesis is not skeptical.
- Many people in fact believe it.

b) The Computational Hypothesis is not skeptical.
- It is just a theory about what constitutes physical particles.
- Other theories about what constitutes observable objects are not skeptical.
- Q: Could reality be fundamentally computational? Maybe not.

c) The Mind-Body Hypothesis is not skeptical.
- Many, perhaps most people already believe this (mind-body dualists).

d) The combination isn't skeptical.

III. Objection: Simulation Is Not Reality

1. Matrix Hypothesis says that there is a computer simulation of physical reality.
2. This does not entail that physical objects exist.
3. Metaphysical Hypothesis entails that physical objects exist.
4. So Matrix Hypothesis does not imply Metaphysical Hypothesis.

Reply:

a. "It is clearly possible that a computational level underlies real physical processes."

b. Two characteristics of computational processes: (i) the structure & complexity of the computation, (ii) how it is implemented.

c. Any adequate computer simulation of reality "will have a rich enough causal structure that it could in principle underlie physics in our world."
d. It does not matter how the computation is implemented.
e. So any adequate computer simulation would constitute a physical reality.

IV. Other Objections

Some beliefs of the brain in the vat that would seem false:

a) "I am outside."
b) "I am in Tucson."
c) "I have hair."
d) "I have friends."

Why these beliefs are really true:

- "I" (in [a]) refers to the brain's virtual body.
- "Tucson" refers to Tucson* (virtual Tucson).
- "hair" refers to hair* (virtual hair).
- "friends" still refers to friends, but the BIV has actual friends (the other brains).

Chalmers' Semantic Theory:

- Two kinds of words/concepts (pp. 24-5):
  - Semantically stable ("neutral") terms and concepts: Retain reference when used by BIV's and normal humans.
    - ☞ Includes: mental, causal, logical, mathematical, and "categorical" concepts. Also, "friend" and "action".
  - Semantically unstable terms and concepts: Shift reference when used by BIV's.
    - ☞ Refer to things they are causally related to.
    - ☞ Includes: "natural kind" concepts, proper names, spatial concepts, the concept "physical".
    - ☞ Putnam's example: Twin Earth
    - ☞ Chalmers' example: Terry and Terry* (16)
- So when the BIV says "physical object," it refers to the virtual physical objects in the simulated world. BIV speaks truthfully in saying "there are physical objects around me," etc.
- Chalmers claims that something along these lines is a consequence of his argument, not presupposed by it (25).

A Possible Objection:

- Maybe "physical", or spatial terms, are semantically stable.
- Chalmers' student Brad Thomson has argued against this.
- Related: Searle rejects the Twin Earth-style thought experiments.

Phil. 2750
Notes #5: A Direct Realist Response to Skepticism

I. Competing Theories of Perception

- Direct realism: Perception gives us direct (non-inferential) knowledge of the physical world.
- Indirect realism: We have inferential knowledge about the physical world, based on the character of our experiences.
- Skepticism: We can't know about the physical world.

II. Review of Dretske's and Klein's Responses to Skepticism

- Skeptic says: "BIV Hypothesis is an alternative to Real World Hypothesis. You can't rule out BIVH. So you don't know RWH."
- Dretske: "I can't rule out BIVH, but I still know RWH. BIVH is irrelevant (because not objectively possible)."
- Klein: "Maybe you can defend the closure principle, but then your claim that I can't rule out BIVH begs the question."

III. What Is Wrong with these Responses

The Courtroom Case

S is on trial for murder. The prosecution offers as evidence the fact that S's blood was found at the scene of the crime along with the victim's blood. They suggest that S got cut while stabbing the victim. The defense offers an alternative hypothesis: S is innocent, and the blood was planted at the crime scene by the police, seeking to frame S.

Dretske: For all we know, the defense is correct, and S was framed by the police. But anyway, we still know S is guilty.

Klein: Either Dretske is right, or the defense attorney's argument just begs the question.

The Scientific Case

Physicist A supports the Copenhagen Interpretation of quantum mechanics. Physicist B supports Bohm's Interpretation. Both interpretations explain all the same data, but they are incompatible with each other.

Dretske: I don't know whether Bohm's theory is right. But I know the Copenhagen theory is right.

Klein: Dretske is right, or Physicist B is begging the question.

IV. A Reformulation of the Argument & the DR Response

1. If E is some evidence and H_1 and H_2 are competing explanations of E, then S is justified in believing H_1 on the basis of E only if S has independent grounds for rejecting H_2. (Preference Principle)
2. The BIV Hypothesis and the Real World Hypothesis are competing explanations of our sensory

experience.

3. So in order to believe RWH on the basis of our sensory experience, we must have independent grounds for rejecting BIVH. (From 1, 2.)
4. We have no such grounds.
5. Therefore, we are not justified in believing RWH on the basis of our sensory experience. (From 3, 4.)
6. Facts about sensory experience are the only justification we might have for RWH.
7. So we're not justified in believing RWH. (From 5, 6.)

Notes:

- This argument escapes Klein's and Dretske's responses, as it should.
- But it only refutes indirect realism.
- (6) is false. We might be foundationally justified in believing RWH, or we might be justified in believing RWH on the basis of facts about the physical world.
- (4) may be false. Facts about the physical world (of which we're directly aware) might be grounds for preferring RWH over BIVH.

V. Objections

(i) Does the DR theory imply that perceptual beliefs are indefeasible?

No. See concept of "prima facie justification": perceptual beliefs are "presumed innocent until proven guilty," i.e., justified as long as there are no positive grounds for doubt.

(ii) Does the DR response 'beg the question'?

Two kinds of responses to skepticism:

- Aggressive response: positive argument, addressed to skeptic, that we have knowledge of the external world.
  - We have not provided one of these.
- Defensive response: response to skeptic's argument that we don't have knowledge of the external world.
  - We have provided one of these.

Phil. 2750
Notes #6: The Problem of Other Minds

I. Problem

- All you observe of other people is their physical behavior.
- Behavior does not logically entail any mental states.
- Hence, how do you know others are not mindless automata (without consciousness)?

II. Some Responses

The Argument by Analogy:

- You observe a correlation between
  a. Your mental states and your resultant behavior.
  b. Things that happen to you and your subsequent mental states.
- You observe other creatures who are outwardly similar to you.
- You infer (by 'analogy') that the same correlations probably hold for them. Hence, that they probably have similar mental states in similar circumstances.

Inference to the Best Explanation:

- You observe people's behavior.
- Hypotheses about their mental states would explain their behavior.
- You know of no other explanations.
- Hence, hypotheses about their mental states are the best available explanations for their behavior.
- Hence, you infer that they probably have mental states of various kinds.

Is It an Innate Capacity?

- Some say that we have an innate capacity to "perceive" others' mental states without reasoning. People's mental states are unconsciously revealed, and perceived by others, through their facial expressions, tone of voice, etc.
- Is this a possible account of our knowledge of other minds?

III. About Heinlein's Story, "They"

- Character claims to have evidence that most other people are mindless automata. Review: what is this evidence? Is it any good?
- What is his evidence for the conspiracy against him? Is his conclusion reasonable? If so, at what point does it become reasonable?

Phil. 2750
Notes #7: Time Travel

I. Issues/Problems about Time Travel

- Some things that commonly happen in time travel stories:
  - Backward causation: When A causes B, even though A occurs after B.
  - Circular causation: A series of events that loops back on itself; A causes itself (directly or indirectly).
  - Self-defeating causal chains: When A prevents itself (or one of its causes) from occurring.
    - ✏ The Grandfather paradox: I decide to go back in time and shoot my grandfather (before my father is conceived). But this would result in my never having been born, so . . .
  - Changing the past: The time traveler causes something to happen differently from the way it "originally" happened.
    - ✏ This seems to involve two time dimensions
- Time travel stories to discuss:
  - "All You Zombies"
  - *Timeline*

  Are they consistent?
- A basic challenge to concept of time travel: Is time travel contradictory? Does it imply that the time traveler *will* (in the future) be in the *past*?
- Two alternative interpretations of "time travel" stories:
  - *The Time Machine story*: In this story, a person "travels back in time".
  - *The Annihilation Machine story:* A crazy person appears with a bunch of beliefs about the future. Later, a person just like him is born in the normal way, grows up, and gets into a machine that annihilates him.
  - The Time Machine story and the Annihilation Machine story involve "the same things happening at the same times". They are different interpretations of the same events (in a sense).
  - Claim: The Annihilation Machine story is the correct interpretation.
    - ✏ Objection: "The Annihilation story is highly improbable."
    - ✏ See why this is confused.

II. Some Ways to Make Time Travel Coherent?

- The branching timeline theory.
  - The "time machine" sends you to an earlier time in a parallel world.
- On the alleged contradiction & the 2 alternative interpretations of "time travel": Distinguish 'personal time' and 'external time'. (Lewis)
  - Personal time: defined by processes internal to the person (mental & physical).
  - External time: defined by external processes.
  - Time travel stories involve a separation between personal & external time: an event is in the 'future' in personal time, but in the 'past' in external time.
  - Is this concept of "personal time" correct? Is there such a time dimension?
- Fixed points and self-supporting causal sequences.

- Something comes out of the "time machine" at the earlier time that causes that same thing to wind up in the position right in front of the time machine at the later time.
  - ✏ The theory is that *whatever* the environment is like, there will be *something* that could come out of the time machine that has that property.
  - ✏ Whatever it is, it will be what comes out.
  - ✏ This is a "fixed point": A fixed point of a function *f* is a value of x for which *f*(x) = x.
  - ✏ Fixed point theorem: A theorem showing that a given function must have a fixed point. Such theorems exist for many functions.
  - ✏ Generally depend on continuity assumptions: *f* should map a space of possible states continuously *into* itself. (See fig. 1.)
  - ✏ The existence of a fixed point depends on the topology of the state space. (See fig. 2.)
- This eliminates "changing the past" and the Grandfather Paradox. It allows circular causation.
- Unclear whether this solution works.

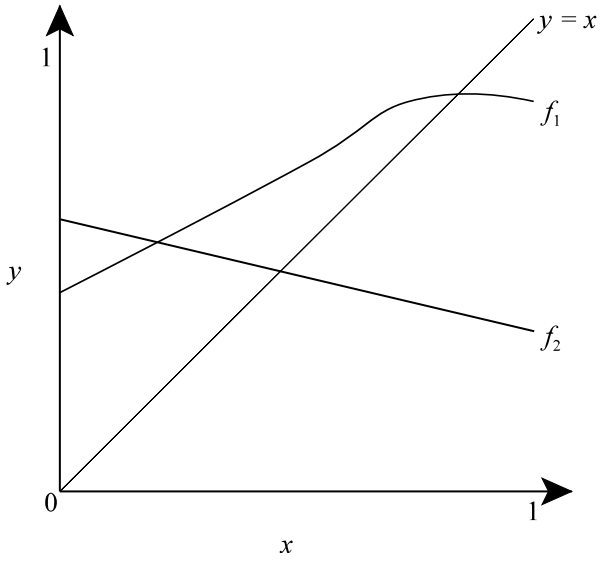


Figure 1. A simple fixed-point theorem: Assume that *f* maps [0,1] continuously into [0,1]. *f* must have a fixed point, for it is impossible to move continuously from x=0 to x=1 without intersecting the x=y line.

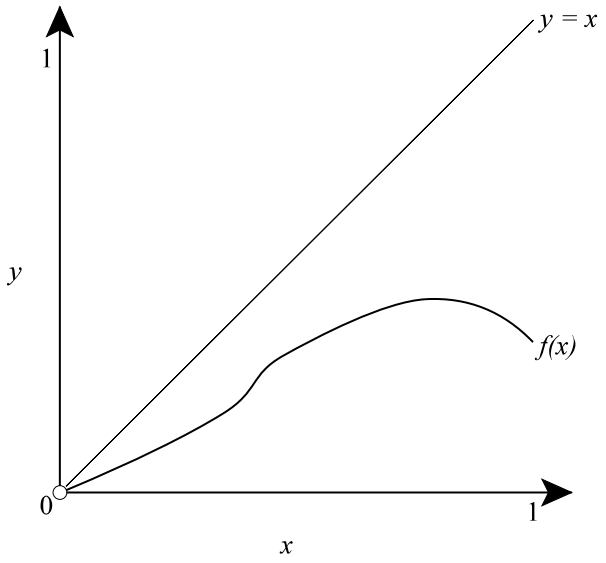


Figure 2. A function with no fixed point: Assume that *f* maps (0,1) continuously into (0,1). *f* may have no fixed point. When the single point x=0 is removed, one can move from the left to the right edge of the graph without touching the x=y line.

Phil 2750
Notes #8: Possible Worlds

I. The Concept of Modality

- Modal concepts: "possible", "necessary", "impossible", "contingent". These are inter-definable:
  P is necessary = ~(Not-P is possible).
  P is impossible = ~(P is possible).
  P is contingent = (P is not necessary & P is not impossible).
- The philosopher's use: possibility = *logical* possibility.
  - Distinguish from: physical possibility, epistemic possibility, feasibility. These are restricted forms of possibility.
- Importance of modality: involved in other concepts:
  - Conditionals ("if A had been the case, B would have been the case")
  - Free will
  - False propositions
  - Meanings of propositions in general
  - Dispositional concepts: Fragility, dangerousness, disgustingness, etc.
- How to understand modal concepts? A standard view:
  - Logically necessary = true in all logically possible worlds.
  - Physically necessary = true in all possible worlds in which the actual laws of nature hold.
  - Epistemically necessary = true in all possible worlds in which our actual knowledge holds.

II. Lewis' Craziness

- There are 'possible worlds.'
- Other possible worlds are worlds, in exactly the same sense that the actual world is one.
- They exist, in exactly the same sense that the actual world exists.
- They each have their own spacetime, disconnected from our spacetime. Like parallel universes, with no way to travel between them & no spatiotemporal relations between worlds.
- 'Actual' just means "pertaining to the world that I'm in."

An argument for the existence of 'possible worlds':

1. Some modal statements are true. (E.g., "I could have had a V8.")
2. Modal statements are best interpreted as assertions about possible worlds, as indicated below:
   "It is possible that p" = "In some possible world, p."
   "It is necessary that p" = "In every possible world, p."
   "It is impossible that p" = "In no possible world does p hold."
3. Therefore, possible worlds exist.

III. Alternative Views

1. Modal expressions are unanalyzable.
   Objection: "This is not an alternative theory at all, but an abstinence from theorizing."
2. 'Possibly, P' = ' "P" is a consistent sentence.'
   Objection: What does "consistent" mean?

a. "consistent" means "could be true."
   Problem: Then the theory is circular.
b. "consistent" means "whose denial cannot be derived from some formal system."
   Problem: From Godel's theorem, for any (consistent) formal system, there are truths of arithmetic that cannot be derived from it. The negation of such a sentence is therefore 'consistent' according to (b). But the negation of a truth of arithmetic is not possible.

3. "Ersatz possible worlds": There are 'possible worlds', but they're really just sets of sentences.
   Objection: this will run into the same problem as (2).

IV. Objections to Possible Worlds

1. Only our own world actually exists.
   Reply: True, but "actually" only means "relative to this world".
2. Realism about p.w.'s is unparsimonious: there are too many entities in your theory.
   Reply:
   - Distinguish (a) qualitative simplicity: reduction in the number of *kinds* of things in a theory, (b) quantitative simplicity: reduction in the number of *instances* of a given kind.
   - My (Lewis') theory has qualitative simplicity.
   - Qualitative simplicity is all that matters.
3. Quine says possible objects are hard to individuate. Not so on my (Lewis') theory, since they are exactly the same sorts of objects as the objects in our world, except that they happen to be in other worlds. Each possible object occupies only its own world.
4. (Not really an objection.) Tell us more about p.w.'s. How many are there? Are there multiple qualitatively indistinguishable p.w.'s?
   Reply: I don't know, and I don't know any way to find out.

V. Methodological/epistemological points:

- We start out with pretheoretical 'opinions.'
- Philosophy should seek a *systematic theory* that *respects (or explains the truth of?) those opinions*.
- Also, we seek qualitatively simpler theories, ceteris paribus.

Phil 2750
Notes #9: Mysteries of Quantum Mechanics

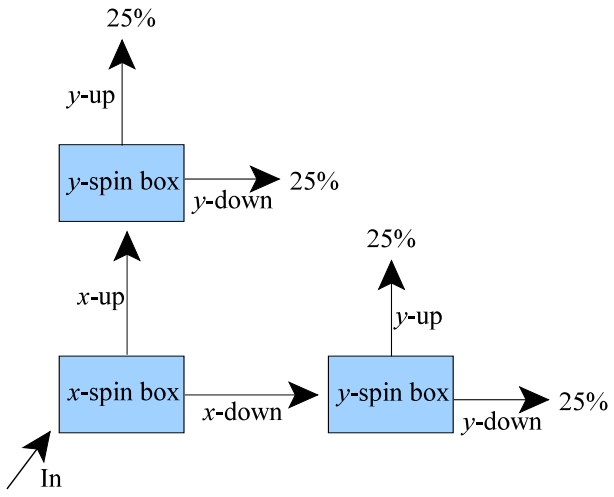
I. Background: Philosophical Interest of Quantum Mechanics

- Some things sometimes said about QM:
  - That it shows that the universe is governed by chance.
  - That it shows that observers create reality.
  - That it converges with Eastern mysticism.
  - That it shows that the Law of Excluded Middle is false, that reality is indeterminate.
    - ✏ LEM: The principle that for any proposition A, either A or ~A.
  - That it shows that all of reality is interconnected.
  - That it shows that maybe there are infinitely many parallel worlds!
- Questions you should ask:
  - Are any of these things true?
  - Why would people say these things?
  - How *could* such things be demonstrated?
- Must understand QM to answer those questions.

II. Electron Spin Mysteries

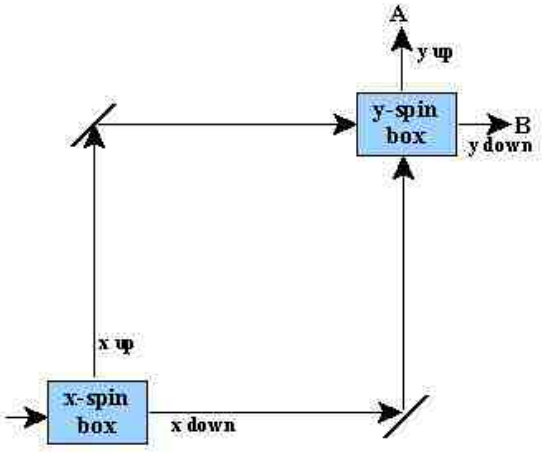
- Electrons have a property called "spin". About spin:
  - An electron has spin in any given direction, e.g., "spin in the *x* direction", "spin in the *y* direction", "spin in the *z* direction". These are distinct.
  - The spin in a given direction can take one of two values: "spin up" (spin +½) and "spin down" (spin -½).
  - Spin affects behavior in a magnetic field. Spin-up electrons are deflected up by a certain amount in a nonuniform magnetic field. Spin down electrons are deflected down by the same amount.
  - Spin in orthogonal directions is completely uncorrelated.

- Measurement:
  - An "*x*-spin box" is a device that measures *x*-spin (spin in the x direction), and sends spin up electrons out in one direction, and spin down electrons out in another direction.
  - Similarly for a "*y*-spin box".
  - Successive measurements of *x*-spin are 100% correlated. Similarly for *y*-spin.
  - But measuring *x*-spin completely *randomizes* *y*-spin, and vice versa (see picture at right).



- A mystery: We feed electrons with various properties into the device at right. Here's what we would expect:

What goes in	What should come out
x-spin up	50% at A, 50% at B
x-spin down	50% at A, 50% at B
y-spin up	50% at A, 50% at B
y-spin down	50% at A, 50% at B



- Here's what actually happens:

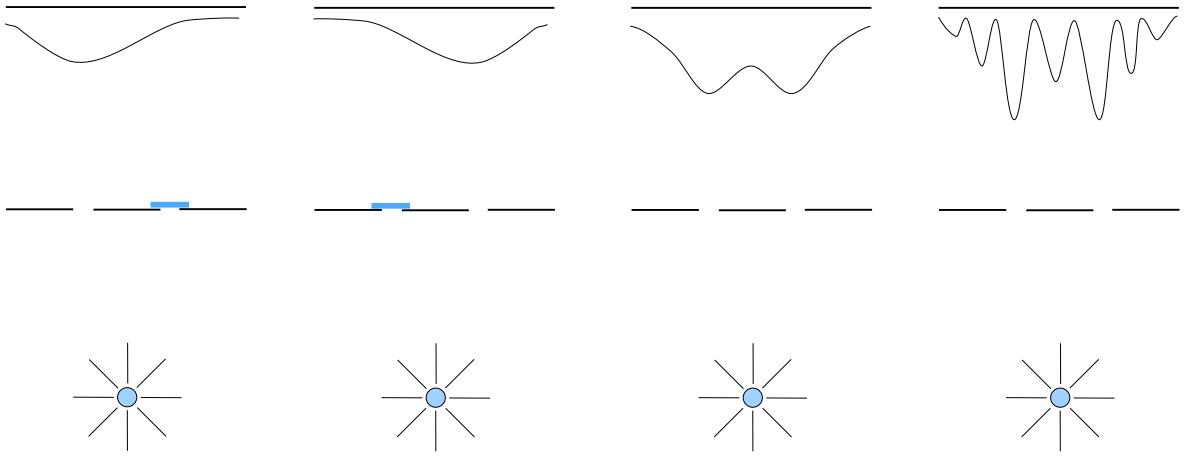
What goes in	What comes out
x-spin up	50% at A, 50% at B
x-spin down	50% at A, 50% at B
y-spin up	100% at A
y-spin down	100% at B

- Superposition (Albert): Say we feed a *y*-up electron into the device. It comes out at A.
  (1) The electron doesn't (simply) take the upper path.
    - If the lower path is blocked, all the electrons coming through are *x*-spin up, and have a 50% chance of coming out at B.
    - Similarly if an electron detector is placed on either path to find out where the electron is.
  (2) The electron doesn't take the lower path.
    - Ditto.
  (3) It doesn't take both paths.
    - If electron detectors are placed, an electron is always found along one path or the other.
  (4) It doesn't take neither path.
    - Ditto.
    - If both paths are blocked, nothing gets through.
  (5) We say the electron is in a *superposition* of both paths, and a superposition of *x*-up and *x*-down.

III. The Double Slit Experiment

- We shoot particles at a wall with 2 slits in it. Behind this wall is a fluorescent screen. We can see where the particles hit the screen. See diagram below.
- If just one slit is open, we get the distribution in (a).
- If the other slit is open, we get the distribution in (b).
- If both slits are open, we *expect* the distribution in (c).
- Instead, what we get is (d).
  - This is an *interference pattern*, explained by wave mechanics.
  - This occurs even if the particles are sent through one at a time.
- If any sort of detectors are placed to determine which slit the particle goes through, the distribution turns into (c).
- Which slit does the particle go through?
  - When measured, the wave/particle acts like a particle, and goes through one slit or the other. No interference pattern.
  - When not measured (with respect to which slit it goes through), it acts like a wave. Parts of

the wave from both slits interfere with each other. The “particle” is in a “superposition” of going through the left slit and going through the right slit.



(a) (b) (c) (d)

Phil. 4400
Notes #10: Quantum Mechanics: The Standard View

I. Background

- QM contains an algorithm for predicting results of measurements.
- The measurement results described above, and many more, are predicted by this algorithm.
- QM is strongly confirmed:
  - These measurement results are very surprising.
  - Classical physics fails to predict them.
  - QM predicts them.
  - This is strong evidence that the algorithm *somehow* reflects the nature of physical reality.
- Q: How should this algorithm be *interpreted*?

II. Basics of the Algorithm

Physical states:

Represented by *state vector*, or *wave function*.

Observables:

These are measurable properties. Represented by *operators* on vectors. Important distinction:

- Properties: This is something that can take on multiple different values (e.g. "position", "charge", "spin").
- States: These are specific values of a property (e.g., "(0,4,5)", "charge of -1", "spin +½").

Determinate states:

- *Eigenvector of an operator:* This is a vector that has a special mathematical relationship (never mind exactly what it is) to the given operator.[1]
  - Most vectors are *not* eigenvectors of a given operator.
  - Each operator has its own, distinct set of eigenvectors.
- When a physical system's state vector is an eigenvector for operator O, then the physical system has a definite value for the property corresponding to O. (We say the system is in an "eigenstate" of that property.)
- Important: When a system is in an eigenstate for a given property, it is generally *not* in an eigenstate for certain other properties.
  - Example: *No* vector is an eigenvector for both *x*-spin and *y*-spin.
  - No vector is an eigenvector for both position and momentum.

What happens when we're not looking:

When physical systems are not being observed/measured, their state vectors evolve in accordance with a *deterministic* law known as the Schrödinger Equation.

What happens when we look:

a) For systems in an eigenstate of O:

---

[1]The mathematical relationship is this: If v_e is an eigenvector of O, then $O(v_e)$ equals a constant times v_e. That is, O fails to change the direction of v_e, and only (at most) changes its length.

- Measurement of O will definitely (100% probability) find the system to be in the corresponding state.
- The state is not disturbed by the measurement.

b) For systems *not* in an eigenstate of O:
- Measurement of O will cause the state vector to jump to an eigenstate of O. The probability of jumping to a given eigenstate is given by a rule known as "the Born rule."[2]
- System will be measured as having the corresponding state.
- Subsequent measurements will have the same result.

III. The Copenhagen Interpretation

1) The wavefunction is complete: there is nothing in physical reality that the wavefunction doesn't represent.

2) Reality is indeterminate: sometimes objects fail to have, e.g., definite locations. Follows from (1).
   - Example: Heisenberg Uncertainty Principle: No physical object ever has a specific location and a specific momentum at the same time.
   - This is *not* an epistemological claim. It is a metaphysical claim.
   - The QM formalism is incapable of even *representing* an object as having such simultaneous pairs of determinate states!

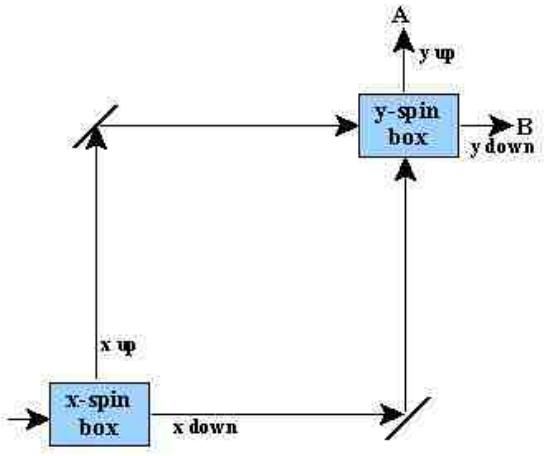
3) "Observation" or "measurement" induces a *physical change* in the system being observed.

4) The world is *indeterministic*: which state a system jumps to on measurement is fundamentally random.

IV. Explaining the Electron Spin Experiment

Suppose y-up electron goes in at lower left:

- As long as no one *looks at* which way the electron exits the *x*-spin box, or it hasn't interacted with a "macroscopic" object, the *x*-spin hasn't been measured yet.
  - Wave function has not collapsed.
  - Electron in a superposition of *x*-up and *x*-down.
  - The *y*-up vector is *identical* (mathematically equal) to a superposition of *x*-up and *x*-down vectors.



[2]This is the rule: suppose we're measuring a certain property. O is the operator corresponding to that property. Suppose v_e is the eigenvector corresponding to some particular value of the property. And suppose v_s is the current state vector of our physical system. Then the probability of the system jumping to the state corresponding to v_e is $|v_e \cdot v_s|$, the absolute value of the inner product of v_s and v_e.

- Also: superposition of upper path and lower path.
- When electron reaches *y*-spin box, it is still *y*-up. Hence, measured to be *y*-up.

Suppose x-up electron goes in:

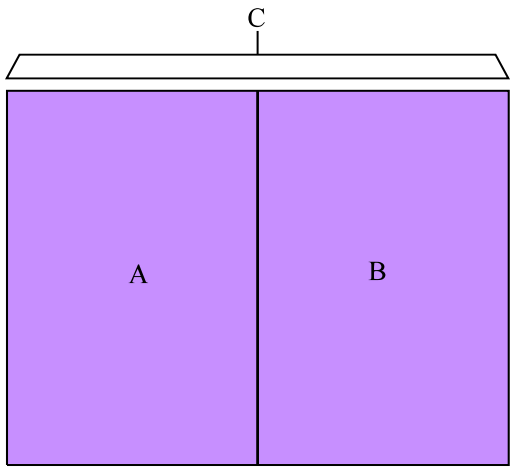
- Takes upper path through device.
- Arrives at *y*-spin box with *x*-up.
- Undergoes wavefunction collapse, with 50% probability of becoming *y*-up, 50% probability of becoming *y*-down.

Suppose y-up electron goes in, and we observe which path it takes:

- The observation of which path it takes induces a wavefunction collapse, with 50% probability of electron becoming definitely on the upper path, 50% probability of becoming definitely on the lower path.
- Subsequently, the electron is *x*-up (say) and on the upper path.

V. Criticisms of Copenhagen

- Entails that observers/measuring devices are governed by different laws from the rest of the universe.
- What constitutes a measurement/observation?
  - When there is a conscious being?
  - When there is an interaction with a "macroscopic" object? How large?
  - Gives rise to the Shrodinger's Cat example. Is the cat enough to collapse the wave function?
- The CI is self-contradictory:
  - Assume an electron is definitely in region C, but not in any definite part of it, i.e., it's in a superposition of all different parts of C. (See diagram.)
  - Let A be the left half of C, and B be the right half of C.
  - CI tells us:
    - ☞ Electron is in (A or B).
    - ☞ Electron isn't in A.
    - ☞ Electron isn't in B.
  - That is a contradiction.



Phil. 4400
Notes #11: Many-Worlds Interpretation of QM

I. The Theory

- When measurements happen:
  - World splits into multiple parallel worlds.
  - Each possible result happens in some world.
- How many worlds?
  - How to explain the probabilities of standard QM: ratio of # of worlds in which a given result happens.
  - The probabilities can take on real-# values.
  - Hence, you need *infinitely many* worlds. (Continuum many, in fact.)
- When does the splitting happen? Two versions of the theory:
  a) Whenever a superposition occurs.
  b) Only when a *measurement* of a system in a superposition is carried out.

II. Advantages of this Theory?

- Preserves determinism.
- Consistent with Locality?
- Might avoid the self-contradiction objection, solve Schrödinger's cat problem?
  - Only if you posit world-splitting whenever superpositions exist. See problems with this below.

III. Criticism of the Theory

- Violates Occam's Razor: posits too many things.
- Not testable: Many worlds are not observable.
- Why don't we bump into the other copies of ourselves?
- When do worlds split?
  a) Superposition version: Problems:
    - *Every* quantum state is a superposition of *something*. Determinacy could only be preserved for some, privileged properties. Choice of preferred properties is arbitrary.
    - Do the parallel worlds affect each other?
      - If no: Then no explanation for interference effects, as in double-slit experiment.
      - If yes: Then no explanation for *lack* of interference effects *after* measurement.

  b) Measurement version: Problems:
    - Still assigns special role to "observers" or "measurements".
    - Still allows superpositions for unmeasured systems. Still subject to self-contradiction objection.

Phil. 4400
Notes #12: Bohm's Interpretation of QM

I. Basic postulates

(a) A physical system consists of particles *and* a pilot wave.

(b) The wave always evolves in accordance with the Schrödinger Equation. No collapse.

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi$$

(c) Particle has a determinate but unknown initial position. The (epistemic) probability of its being at a location is proportional to the square of the amplitude of the wave function at that location. $\rho = |\Psi|^2$

(d) The wave causes the particle to move in a specific way. The particle gets carried along with the flow of the amplitude of the wave function, according to the equation below. (It moves in the direction of the gradient of the wave function.) The equation of motion:

$$\frac{dQ_k}{dt} = \frac{\hbar}{m_k}\mathrm{Im}\frac{\Psi^*\nabla_k\Psi}{\Psi^*\Psi}(Q_1,\cdots,Q_N)$$

(e) For a system of many particles, there is a *single* wave, occupying a many-dimensional configuration space. The equation in (d) determines the change in the position of the *system* in *configuration space*.

Configuration space:

- A mathematical "space" that a system occupies, with three dimensions for each particle.
- The system occupies a point in that space. The location of the system in the configuration space reflects the locations of all the particles in physical space.
- Example: Consider two particles, A [located at (1, 3, -4)] and B [located at (-1, 2, 0)]. The 2-particle system occupies the point (1, 3, -4, -1, 2, 0) in the 6-dimensional configuration space.

II. Interesting features

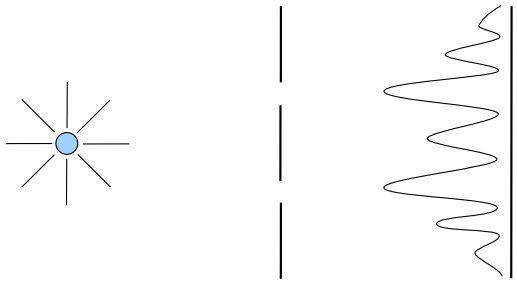
- Postulates (c) and (d) entail that the probability distribution for particle positions at any *later* time will *also* be proportional to Ψ^2. (The equation of motion is specifically cooked up to achieve this result.)
- The theory is deterministic. But indeterministic variants can be developed. (As suggested by Bohm & Hiley in *The Undivided Universe*.)
- The theory gives the standard empirical predictions of quantum mechanics.
  - One exception: If a sufficiently precise collapse theory is given, it is possible in theory (but extremely difficult) to test for wave function collapses. Bohm predicts that no such collapse will be found.
- The theory is nonlocal. Instantaneous action at a distance is possible. Bohm says everything is interconnected.
- All properties other than position are "contextual".
  Contextual properties: Properties that depend on a relationship of the object to its environment (esp. the experimental apparatus used for "measuring" them).

- Ex.: Outcome of a spin measurement depends on orientation of the measuring device.

III. How does the theory deliver QM phenomena?

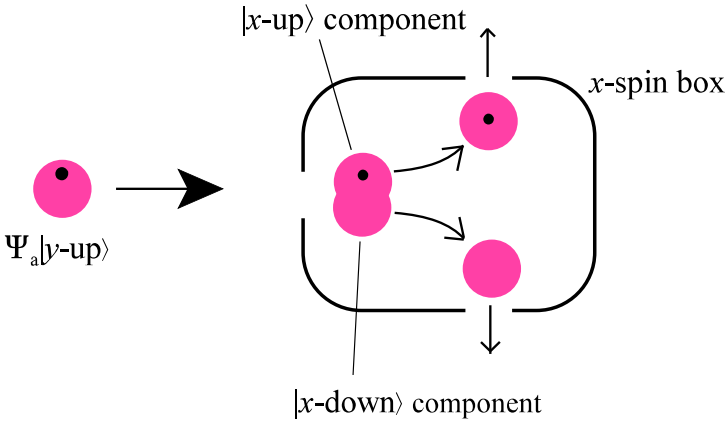
A) The double slit experiment

- Pilot wave goes through both slits, producing interference.
- Particle goes through one slit or the other, depending on its initial position.
- The equation of motion [(d) above] implies that the particle will be carried away from areas where the amplitude of the wave is lowest. Hence the observed interference pattern.
- If one slit is blocked, the pilot wave only goes through the other. No interference.



B) A spin measurement

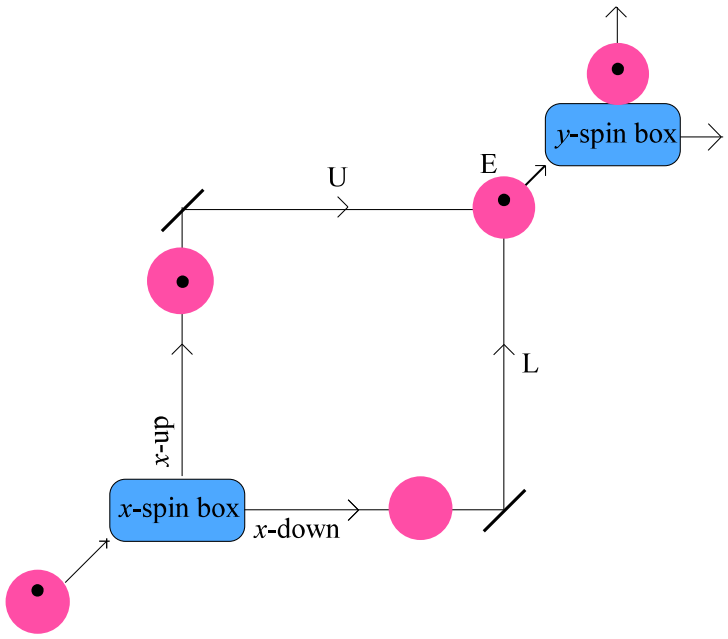
- Electron starts out with $|y\text{-up}\rangle$ wave function, is fed into *x*-spin box.
- The wave function evolves: The $|x\text{-up}\rangle$ component of the wave function moves towards the aperture indicating "spin up", while the $|x\text{-down}\rangle$ component moves the opposite way.
- Electron carried along with the wave function. Suppose electron starts in upper half of the wave. Then it will move up.
  - As the "*x*-up" and "*x*-down" components move apart, the electron winds up in a region where the "*x*-down" component is absent *first*.
  - It is subsequently carried along by the "*x*-up" part of the wave function.
- *Note:* No collapse. The $|x\text{-down}\rangle$ part of the wave function still exists.
- Why will electron subsequently be measured as *x*-up, with 100% probability?
  - $|x\text{-down}\rangle$ component is no longer in the region where the electron is located.
  - Electron's motion determined by the wave function at its current location.
  - This explains the *apparent* collapse.
- But if the *x*-down component is somehow brought back to where the electron is, it can then affect the electron's behavior.



C) The mysterious two-path experiment

- Electron takes one path or the other.

- The wave splits in two and takes both paths.
  - The |*x*-up⟩ component takes the U path.
  - The |*x*-down⟩ component takes the L path.
- The wave components recombine at point E, creating a |*y*-up⟩ wave function once again. (See diagram.)



D) "Effective" collapses

- Measurement brings about "effective collapse": the wave function does not actually collapse, but system acts as if it did.
  - Components of the wave function corresponding to different possible measurement outcomes (outcomes that would have occurred had the initial position been different) still exist.
  - But these components separate into different places in configuration space.
  - Only the wave function components in the vicinity of *the system's current position* affect its motion.
- The non-existence of collapses is *in principle* detectable. After a measurement:
  - Different components of the wave function (corresponding to different possible measurement outcomes) must be recombined in configuration space.
  - An 'interference effect' (as in the double slit experiment, or the 2-path experiment above) would occur, according to Bohm.
  - Collapse theories predict no interference effect.
- This is *in practice* unfeasible. Why:
  - Recombination of wave function components implies:

    The system is at a place in configuration space, such that it would have been at that same place, if one of the other measurement outcomes had occurred.

  - That means: *Every particle* in the system is where it would have been (in physical space),

had another measurement outcome occurred.
- That means: Every trace of the measurement outcome, in the position of any particle, has been erased.

IV. Advantages of Bohm

1. *Uniform dynamics:*
   - Wave function always evolves in the same way. (The collapse postulate is bogus.)
   - Measuring devices/observers governed by same laws as the rest of physical reality.
2. *Logical coherence:* Cats are either alive or dead, not in a 'superposition' of alive and dead.
3. *Precision:* Copenhagen interpretation requires a vague concept of "measurement," or "macroscopic" objects.
4. *Determinism,* if you consider that an advantage.

Lesson:

None of the things usually said to be supported by QM need be accepted. Except one: Bohm's theory *does* imply that everything is interconnected.

V. Objections to Bohm

1. *Conflicts with Special Relativity.*
   - Bohm's theory is nonlocal: Instantaneous action at a distance is possible.
   - This conflicts with Special Relativity:
     - In SR, the time-order of events outside each other's light cones is relative to a reference frame.
     - Bohm's theory requires such events to have a determinate time order (it matters which event affects which).
     - So Bohm requires a preferred reference frame.
   - But we can *not* use the nonlocality to send signals.
   - We also cannot *identify* the preferred reference frame.
2. *A positivist objection:* Bohm's theory entails the existence of undetectable facts (as just noted).
3. *Copenhagen got here first.* Bohm's theory lacks novel predictions.
4. *The conspiracy of silence objection:* Isn't it bizarre how the world conspires to prevent us from detecting (a) the preferred reference frame, (b) the truth of determinism, (c) the lack of collapses? I.e., things are set up exactly to make it look like orthodox QM is true?
5. *The probabilistic postulate,* $\rho = |\Psi|^2$*, is ad hoc.* Why should the epistemic probability be that?
6. *Technical objection:* Difficult to come up with a Bohmian version of relativistic quantum field theory.

Phil. 4400
Notes #13: The Doomsday Argument

I. Doomsday Possibilities

- Perhaps mankind will soon go extinct. How could this happen?
  1. Nuclear war.
  2. Environmental destruction.
  3. Asteroid impact.
  4. Disease, genetically engineered or otherwise.
  5. Misuse of nanotechnology.
  6. We're living in a simulation and it gets shut down.
  7. Badly programmed superintelligence.
  8. Physics experiments.
  9. Something unforeseen.
- Consider two hypotheses:
  (S) *Doom Soon:* Mankind will soon go extinct. This could happen from any of the above causes.
  (L) *Doom Late:* Mankind will survive far into the future. Maybe humans will figure out how to avoid all those threats.
- Other possibilities: There are infinitely many hypotheses as to when the species will go extinct. We consider just these two for the sake of simplicity.
- Mankind will go extinct. The question is when.

II. The Doomsday Argument

A. The Self-Sampling Assumption

- You should treat yourself as a random selection from all observers.
- Example: 100 people, including you, are drugged and kidnaped. Each is taken to a hotel room. 90 of the rooms are painted blue on the outside, 10 are red. Q: Given this information, what is the probability that your room is blue?

B. Another Example

- You are drugged, kidnaped, and taken to a hotel room. You were either taken to a small hotel (with 10 rooms), or a large hotel (with 100 rooms), depending on the toss of a fair coin. Each hotel has its rooms labeled with integers, starting from 1 (room #1, room #2, etc.).
  Q1: Given just this information, what is the probability that you are in the small hotel?
  Q2: Suppose you learn that you are in room #7. *Now* what is the probability that you are in the small hotel?
- Bayes' Theorem: $P(h \mid e) = \dfrac{P(h) \cdot P(e \mid h)}{P(h) \cdot P(e \mid h) + P(\sim h) \cdot P(e \mid \sim h)}$
- In this example, let:
  S = You are in the Small hotel.
  L = You are in the Large hotel.
  E = Your room number is 7.

- Applying Bayes' Theorem:

$$P(S \mid E) = \frac{P(S) \cdot P(E \mid S)}{P(S) \cdot P(E \mid S) + P(\sim S) \cdot P(E \mid \sim S)}$$

P(S) = .5 P(~S) = .5
P(E|S) = 1/10 P(E|~S) = P(E|L) = 1/100

Hence:

$$P(S \mid E) = \frac{(.5)(.1)}{(.5)(.1) + (.5)(.01)} = \frac{.05}{.05 + .005} \approx .91$$

C. The Doomsday Analogy

- You find yourself living as a member of the human species. It is either a short-lived species (per "Doom Soon") or a long-lived species (per "Doom Late").
  Q1: What is the initial probability of Doom Soon?
  Q2: You learn that you are approximately the 60 billionth human born. Now what is the probability of Doom Soon?
- Let
  S = Doom Soon. Suppose this means the total # of humans ever to have lived will be ~100 billion.
  L = Doom Late. Suppose this means the total # of humans ever to have lived will be ~1 trillion.
  E = Your birth rank is ~60 billion.
- What is P(S)? Hard to calculate. Seems like a pretty good chance, somewhere between 10% and 90%. Let's leave it as an unknown, *s*.
- What is P(S|E)? By Bayes' Theorem,

$$P(S \mid E) = \frac{P(S) \cdot P(E \mid S)}{P(S) \cdot P(E \mid S) + P(\sim S) \cdot P(E \mid \sim S)}$$

P(S) = *s* P(~S) = 1 - *s*

$$P(E \mid S) = \frac{1}{100 \text{ billion}} \qquad P(E \mid \sim S) = P(E \mid L) = \frac{1}{1 \text{ trillion}}$$

Hence:

$$P(S \mid E) = \frac{(s)\left(\frac{1}{100 \text{ billion}}\right)}{(s)\left(\frac{1}{100 \text{ billion}}\right) + (1-s)\left(\frac{1}{1 \text{ trillion}}\right)}$$

Multiplying numerator and denominator by 1 trillion gives us:

$$P(S \mid E) = \frac{10s}{10s + (1-s)} = \frac{10s}{9s+1}$$

- Try plugging in some hypothetical values for *s*:

When *s* = 0: $P(S \mid E) = \frac{10(0)}{9(0)+1} = 0$

When $s = 1$: $P(S \mid E) = \frac{10(1)}{9(1)+1} = 1$

When $s = .1$: $P(S \mid E) = \frac{10(.1)}{9(.1)+1} \approx .53$

When $s = .5$: $P(S \mid E) = \frac{10(.5)}{9(.5)+1} \approx .91$

- Conclusion: E supports S for all values of *s* between 0 and 1.
  - The support is stronger, the larger we make the population in Doom Late. If Doom Late has 100 trillion people, then we get:

$$P(S \mid E) = \frac{1000s}{999s+1}$$

When $s = .5$: $P(S \mid E) = \frac{1000(.5)}{999(.5)+1} \approx .999$

III. Criticisms of the Doomsday Argument

A. A disanalogy between IIB and IIC:

- In the analogy in IIB:
  - I already exist before the hotel is chosen. Which of {S,L} is the case only affects *where* I am. If the coin flip came up the other way, I would just be somewhere else.
  - If the Large hotel had been chosen, I would probably now have a higher room #.
- In IIC,
  - Whichever of {S,L} is the case, I would *not* be somewhere else if the other one were the case.
  - It is *false* that if L were the case, I would probably now have had a higher birth order: Everything up to now would be exactly as it is.

B. Whether S or L were the case, **someone** *would have birth order 60 billion (or whatever my birth order is).*

- Therefore, the fact that someone has birth order 60 billion does not support S over L or vice versa.
  - If $P(E|H_1) = P(E|H_2)$, then E supports H_1 if and only if E supports H_2.
  - In this case, P(Someone has birth order 60 billion | S) = 1 = P(Someone has birth order 60 billion | L).
- Surely the fact that that someone is *me*, rather than someone else, is not relevant either.
  - S and L do not make any different predictions about how likely it is that the person in question would "be me".

C. My existence is evidence for L

- Intuitive idea:
  - *Given that I exist,* S makes it ten times more likely that my birth order would be < 100 billion than L does.
  - So "My birth order is < 100 billion" is evidence for S over L.
  - But L makes it ten times more likely that I would exist at all than S does, because in L there would be ten times as many people.

- So my existence is evidence for L.
- These two pieces of evidence exactly counterbalance one another.
- Thus, "I exist and have birth order < 100 billion" is overall neutral between S and L.

• The probability calculations:

Let B = That I exist at some time in the history of the species.

$$P(S \mid E \& B) = \frac{P(S) \cdot P(E \& B \mid S)}{P(S) \cdot P(E \& B \mid S) + P(\sim S) \cdot P(E \& B \mid \sim S)}$$

P(S) = s

P(~S) = 1 - s

P(B|~S) = 10·P(B|S)

$$P(E \& B \mid S) = P(B \mid S) \cdot P(E \mid B \& S) = P(B \mid S) \cdot \frac{1}{100 \text{ billion}}$$

$$P(E \& B \mid \sim S) = P(B \mid \sim S) \cdot P(E \mid B \& \sim S) = 10 \cdot P(B \mid S) \cdot \frac{1}{1 \text{ trillion}} = P(B \mid S) \cdot \frac{1}{100 \text{ billion}}$$

Thus, plugging those things into the first equation:

$$P(S \mid E \& B) = \frac{s \cdot P(B \mid S) \cdot \frac{1}{100 \text{ billion}}}{s \cdot P(B \mid S) \cdot \frac{1}{100 \text{ billion}} + (1-s) \cdot P(B \mid S) \cdot \frac{1}{100 \text{ billion}}} = \frac{s}{s + (1-s)} = s$$

Phil. 2750
Notes #13: Artificial Intelligence

I. The Question of Artificial Intelligence

- Artificial beings frequently represented as conscious, intelligent, etc.
- Is this realistic? Could computers have minds?
  - Our minds include: Beliefs, desires, feelings, sensations, perceptions, memories. These are "mental states".
  - Mental states have two interesting characteristics:
    a) Intentionality: The property of representing the world.
    b) Qualia: What mental states feel like to the subject. A.k.a. "subjective character" of experience.
  - Would robots or computers have these characteristics?
- Searle's presentation:
  - Not the question:
    a) "Could a machine think?" Humans are "machines", and we can think.
    b) "Could a computer think?" The human brain is a "computer".
    c) "Could an *artificial* machine think?" Also uninteresting. An artificial being that was intrinsically just like us could think.
  - **Strong AI** claims: A computer could think *solely in virtue of* running an appropriate program.
    Related idea: **The Turing Test**:
    > A computer and a person are put in different rooms. An expert is then brought in and allowed to communicate with the two by terminal. If the expert is unable to tell which is the human and which is the computer, then the computer passes the test for being conscious.
- Possible misunderstandings:
  - Strong AI is not merely claiming that a computer *might* be able to think.
  - Also not claiming that a computer could *probably* think, or that we'd have *evidence* that it could think, if it passed the Turing Test.
  - The claim is that the computer *ipso facto* would be thinking, because that's all there is to thinking.
- Strong AI stems from "functionalism": The theory that mental states are just "functional states."
  - Functional state: A state that is defined purely in terms of causes and effects.
  - According to functionalism, having a mind is just having internal states with the right network of causal connections with each other.
  - Thus, the mind is like the program run by the brain.
  - This is the dominant contemporary view in philosophy of mind, cognitive science, & AI research.

II. The Chinese Room Thought Experiment

- Assume: a computer exists that passes the Turing Test, communicating in Chinese.
- I am put in a room with a copy of the program. By blindly following the rules in the program, I also pass the Turing Test.
- This does not give me an understanding of Chinese.

• Therefore, it wouldn't give any other computer an understanding of Chinese.

III. Searle's Argument

1. Programs are defined purely syntactically.
2. Minds have semantics (meanings).
3. Syntax is not sufficient for semantics.
4. Therefore, no program is sufficient for having a mind.

• Searle says the thought experiment just illustrates this.

IV. Objections & Replies

A. The Systems Reply:
- "You're just the CPU. The CPU doesn't understand Chinese; it's the whole system."
- Searle says: I memorize the rulebook and work out in an open field.

B. The Robot Reply:
- What if we put the computer into a robot with artificial sense organs, etc., and let it interact with the world?
- Searle says: You could similarly put me in the head of a giant robot, etc.

C. The Brain-Simulator Reply:
- What if we made a computer that simulated the actual sequence of neuron firings in a normal brain?
- Searle says: You could have me simulate the same sequence, using water pipes, etc.

D. The Combination Reply:
- What if we combined the previous three replies? Wouldn't we attribute mental states to a robot that acted just like us, with the same structure as our brains?
- Searle says:
  ☞ Combine Searle's previous counter-replies.
  ☞ We wouldn't attribute mental states if we knew its behavior was the result of a formal program.
    • Searle seems to be saying: (1) We attribute mental states to others as the best explanation of their behavior, but (2) If we knew the robot was just running a program, then attributing mental states to it *wouldn't* be the best explanation of its behavior.

E. The Other Minds Reply:
- But how do you know that other *people* are conscious?
- Searle: This misses the point. The question is what consciousness *is*, not how we know about it.
- Comment: The reply may be confusing an epistemological claim with a metaphysical claim.
  Epistemological claim: X would be evidence for Y.
  Metaphysical claim: X is sufficient for Y; X is just what Y consists in.

F. The Many Mansions Reply:
- Someday, when we find out *whatever* causes consciousness, we'll be able to artificially

produce it.
- Searle: This "trivializes" strong AI by defining it as whatever artificially produces consciousness.

V. Searle's View of the Mind

- Intentionality is a distinctive, non-reducible property of minds.
- Brains cause it.
- The material of which the brain is made has the relevant causal powers.
- Some other material *might* also have those powers. Or it might not.
  - In any case, *instantiating a program* is not what produces consciousness.

VI. Searle vs. Fodor

• ***Fodor:***
- Fodor is a functionalist.
- But to be conscious, one must have *the right kind* of causal connections between one's internal states.
- We don't know what the right kind is.
- But it can't be mediated by a little man in your 'head'.
- This isn't so strange, because a similar situation obtains with other concepts:
  • Perception requires the right kind of causal connection between object and perceiver.
  • "Breaking" an object requires causing it to be broken, in the right sort of way.
  • In neither of these cases may the causal connection be mediated by a little man.

• ***Searle:***
- No kind of causal connection between a symbol and egg foo yung is sufficient for the agent to understand what the symbol means. (Comment: Seems to misunderstand Fodor's claim?)
- Elsewhere, Searle says: Fodor's claim is so implausible--he seems to be saying that by virtue of the fact that the man in the Chinese room is intentionally trying to implement the program, he thereby can't really implement it.

Phil. 2750
Notes #14: The Mind/Body Problem

Q: What is the relationship between the mind and the body?

I. Theories of the Mind & Body

A. Physicalism: Only physical phenomena exist. Two versions:

1. *Eliminative*: There are no mental phenomena.
2. *Reductive*: Mental phenomena are reducible to physical phenomena (e.g., brain states). Versions of this:
   - Mental states are brain states.
   - Behaviorism: mental states are dispositions toward physical behavior.
   - Functionalism: mental states are functional states: states that are defined by systems of causal relations with each other and the environment. (Similar to behaviorism, but more complicated.)

B. Dualism: There are 2 kinds of things, mental & physical. Two versions:

1. *Substance*: Two kinds of substances, minds and bodies.
   Substance (in metaphysics): Something that is capable of independent existence; not a modification or attribute of something else.
2. *Property*: One kind of substance, but 2 kinds of properties. Often associated with:
   The Theory of Emergence (Emergentism): Sometimes, "the whole is greater than the sum of the parts."
   - Certain complex arrangements of matter cause novel properties to emerge that would not have been expected on the basis of studying simpler arrangements.
   - There are "emergent laws" of nature that operate specifically when certain complex arrangements of matter arise. These complex objects are *not* fully explained by the laws that explain the properties and behavior of simpler systems.

C. Idealism: Only mental phenomena exist.

Almost no one holds this.

II. Why Is the Mind/Body Relation Puzzling?

1. People are made of atoms.
2. Atoms are purely physical, with no mental properties.
3. Consciousness can't be deduced/explained from purely physical properties.
4. People are conscious.
5. The characteristics of a system can be deduced/explained from the properties and arrangement of its parts.

• Those five propositions are jointly inconsistent.
• Each one seems to be true. (Arguments discussed below.)
• Each solution to the mind/body problem denies one of these:
  Eliminative materialism: Denies #4.

Reductive materialism: Denies #3.
Substance dualism: Denies #1.
Emergentism (& property dualism): Denies #5.
Panpsychism: Denies #2.
Idealism: Denies #1.

III. Arguments for (1) and Against Substance Dualism

- Appeal to Occam's Razor
- No scientific explanation for origin of non-material minds.
- Problem of how mind & body interact.

IV. For (3) and Against Reductive Materialism

- Thought Experiment:
  Mary is raised indoors, in an entirely black-and-white environment. She learns all the facts about brain science in this black-and-white environment. One day, Mary leaves her room and for the first time, looks up at the sky and sees something blue. Does Mary learn anything new?

1. When Mary leaves her room, she learns something new: what it is like to see blue.
2. Mary does not learn any new physical information.
3. Therefore, what she learns is a non-physical fact.

Phil. 2750
Notes #15: The Problem of Personal Identity

I. Basic Ideas

- Identity: Two senses:
  - Qualitative identity: A.k.a. indiscernibility, indistinguishability. Having the same qualities.
  - Numerical identity: The relation that everything bears to itself and to nothing else.
- We will be concerned with numerical identity.
- Basic question:
  - Under what conditions is *x* the same person as *y*?
  - Or: Under what conditions are two person-stages stages of the same person?
  - *For example:* When you step into the teletransporter, what happens?
    a) You get transported to Mars (or wherever)?
    b) You are destroyed; then, someone who looks just like you is created somewhere else?
- Some principles about identity:
  1. ***Transitivity***: If x=y and y=z, then x=z.
  2. ***Symmetry***: If x=y then y=x.
  3. ***Leibniz' Law*** (the indiscernibility of identicals): If x=y, then for any feature F, x has F if and only if y has F.
  4. ***Uniqueness***: *Two* people cannot be identical with each other.
  5. ***Instrinsicness:*** Identity is intrinsic; whether x=y does not depend on things going on outside x and y.

II. Some Theories about Personal Identity

A. The body theory: A person is his body. x=y iff x and y have the same body.
Objection: Brain transplant case.

B. The brain theory: A person is his brain.
Objection: Mind transplant case?

C. The soul theory: A human being has an immaterial component, the "soul". A person is his soul.
Objection: Do souls exist?

D. The memory theory: Memory determines identity. Some versions of this:
- x is a later stage of y iff x has all the same memories as y.

Objections: Cases of forgetting, amnesia.

- x is a later stage of y iff x remembers y.

  Objection: Violates transitivity (Reid's objection).

- x is a later stage of y iff there is a chain of stages from x to y, where each stage remembers the one before it.

  Objection: Amnesia case. Violates Uniqueness.

E. The spatiotemporal continuity theory: x is a later stage of y iff there is a spatiotemporally continuous series of person-stages connecting x to y.
Objections: Cases of radical psychological change. Reincarnation. Violates Uniqueness.

F. The closest-continuer theory: x is a later stage of y iff: x does sufficiently well on a set of criteria for identity, and no other thing existing at the same time as x does better. The criteria may include: spatiotemporal continuity, psychological continuity, sharing memories, sharing psychological traits, sharing body material.
Objection: Violates Intrinsicness.

G. The Skeptical Theory: There are no objective facts about personal identity. It is a semantic question, a matter of convention.
Objection:

- "Will I have such-and-such future experience?" isn't a semantic question.
- What are the implications for practical (prudential) reasoning?
- Could you make yourself immortal by adopting suitable conventions?

III. What Do We Believe Ourselves to Be? Back to the Soul Theory

A. Against Purely Qualitative Theories

- Qualitative theory: Any theory according to which a continuer of me-now merely has to have certain qualitative characteristics. More generally: "x is a continuer of y iff xRy, where R is any relation such that more than one thing can bear it to the same person.
- Examples: Memory, spatiotemporal continuity
- All such theories violate Uniqueness.
  - Suppose xRy and zRy, where x and z are two distinct persons.
  - Then, according to the theory, x=y and z=y.
  - By Transitivity, x=z. But this is not the case.

B. Against the Stipulative Qualitative Theory

- This is like the Pure Qualitative approach, but with the stipulation that there may not be a second object that bears R to y. I.e.: "x is a continuer of y iff xRy,

and there is no other z such that zRy."
- Violates Intrinsicness.

Intermediate Conclusion: Personal identity over time is a matter of some *thing* continuing to be present, not a matter of having certain *qualities* in common with the original person. What could this thing be?

C. Against the Body Theory

- As above, mind transplant, reincarnation cases.
- Bodily identity is conventional.

D. Conclusion: We Think of Ourselves as Souls

- How strong is the objection to this?

Phil. 2750
Notes #16: Free Will

I. Background to the Free Will Issue

Free Will: Two important components:

a. Self-Control: A free agent controls his own actions.

b. Alternate Possibilities: A free agent is able to do otherwise. Often held to conflict with determinism.

Determinism: Two formulations:

a. Every event has a sufficient cause.

b. At any point in time, there is a unique future course of events that is consistent with the state of the world at that time and the laws of nature.

Traditional positions on the free will issue:

a. *Libertarianism:* FW + not-deteterminism.

b. *Hard Determinism:* No FW + determinism.

c. *Soft Determinism (compatibilism):* FW + determinism.

The importance of FW:

FW is presupposed in

- Emotions: blame, gratitude, remorse, praise, respect, pride.
- Deliberation.
- Perhaps reasoning (see IIIC).
- Punishment/reward.
- "Should" judgements/prescription/judgements of right/wrong.

No remotely normal life is possible to the FW denier.

II. Side Issue: Predictability

- Q: Could there be a device that predicts everything I do?
- Some might think this bears on determinism, because of this argument:
  1. If determinism is true, then everything is predictable (in principle).
  2. If so, then there could (in principle) be a device that correctly predicts all of my actions.
  3. There could not be such a device, for I could easily falsify its predictions.
  4. So determinism is false.
- This argument probably confuses determinism with predictability, in premise (1).
- Could there be a device that correctly predicts all your actions?
  - Its prediction would affect what you're going to do.
  - This doesn't necessarily make it impossible. The device could take into account the effect of its prediction.
    - It could calculate a "fixed point": a prediction such that, if the device makes that prediction,

your mental state is altered in such a way as to make the prediction come true.
- You might be perverse, disposed to do the opposite of what the device says.
  • Maybe the device predicts in another language.
  • Maybe the device lies.
  • Maybe the device keeps its prediction to itself.

• Predictability & quantum mechanics
- Does the Heisenberg Uncertainty principle entail unpredictability?
- Confusion: The HUP is not about *uncertainty* in the standard interpretation; it is about *indeterminacy*.
- Knowing the position & momentum of a particle at once is irrelevant. In QM, behavior is determined by the wave function.
- There are both deterministic and indeterministic interpretations of QM.

III. Is Determinism True?

A. Evidence from physics:
• Humans are made of matter. Matter must obey the laws of physics.
• Classical physics: deterministic.
• As noted, there are both deterministic and indeterministic interpretations of QM.
• Conclusion: physics does not settle whether determinism is true.

B. Other sciences:
• Sometimes said that psychology, social science, etc. supports determinism.
• This is sheer bluff.

C. Main argument against determinism:
1. We have free will. (See below.)
2. The existence of free will is incompatible with determinism. (See below.)
3. So determinism is false.

III. Do We Have Free Will?

A. Against FW:
1. If our actions are determined, then we lack free will, for:
   a. If our acts are determined, then we don't have alternate possibilities.
   b. FW requires alternate possibilities.
2. If our actions are *not* determined, then we lack free will, for:
   a. If our actions are not determined, then they are random.
   b. If our actions are random, then we do not control them.
   c. FW requires self-control.
3. Therefore, we lack free will.

B. For FW: introspection. Some say we can tell by introspection that our choices are free.
- Introspection reveals that we make choices, but the question is *not* whether we *make choices*. The question is whether we have genuine alternate possibilities.
- *Objection #1:* FW requires alternate possibilities. How can introspection reveal the existence of *non-actual possibilities*?

- This is inconclusive. Not obvious that introspection could not reveal non-actual possibilities. Particularly if this is construed as a negative condition, e.g., the absence of determining causes.
- *Objection #2:* Some admit that we have an experience of freedom, but say that this experience may be an "illusion." How can we be sure it isn't?
  - *Reply:* How is this any different from any other philosophical skeptic argument?

C. For FW: self-refutation argument.

The denial of free will is self-refuting, for one of these reasons:

- Because reasoning involves a kind of deliberation. Deliberation presupposes FW.
- Because determinism implies that we do not believe anything *because* it is true or because of the stated reasons for it.
- Because reasoning/discourse presupposes norms of rationality.
  a. If the determinist rejects such norms, he is in a self-defeating position.
  b. If the determinist accepts rationality norms, he is also in a self-defeating position.

An argument along these lines:

1) We should believe only what is true. (Premise, norm of rational discourse.)
2) "S should do A" implies "S can do A". (Premise.)
3) We can believe only what is true. (From 1, 2.)
4) If determinism is true, then S can do A only if S does A. (Premise, def. of determinism.)
5) Therefore, if determinism is true, then we believe only what is true. (From 3, 4.)
6) I believe I have free will. (Premise.)
7) Therefore, if determinism is true, then it is true that I have free will. (From 5, 6.)

Objections to the self-refutation argument:

a. Premise 1 "begs the question". Or: (1) is false because determinism is true.
b. (1) is false. We should believe only what is *justified*.
c. (1) is false because we can't control our beliefs.
d. The argument contains an equivocation, because "should" can be either the moral "should" or the epistemic "should."

IV. Is FW Compatible with Determinism?

A. Yes:

1. FW is incompatible with *in*determinism (see IIIA).
2. FW is possible.
3. So it must be compatible with determinism.

B. Yes:

1. Proposed accounts of freedom:
   • Absence of constraints.
   • Actions caused by one's beliefs, desires, values.
   • "can" = "would if one tried"/"would if one wanted to"
   • Higher order-desires/evaluations: S controls his desires, by his second-order desires.
2. All these things are compatible with determinism.
3. So FW is compatible with determinism.

C. No: The Consequence Argument:

1. If Determinism, then my actions are consequences of events in the remote past + the laws of nature.
2. I have no choice about the remote past.
3. I have no choice about the laws of nature.
4. If I have no choice about *x*, then I have no choice about necessary consequences of *x*.
5. Therefore, if Determinism, then I have no choice about my actions.

D. No: A similar argument:

1. If, in order for me to do A, something would have to have happened in the past that did not in fact happen, then I cannot do A.
2. If determinism is true, then in order for me to do other than what I actually do, something would have to have happened in the past that did not in fact happen.
3. Therefore, if determinism is true, then I cannot do other than what I actually do.

V. Conclusions

- Soft determinism is false.
- Hard determinism is false.
- How is libertarian FW possible?
  - Agent-causal theory?

Phil. 2750
Notes #17: Gödel's Theorem and Free Will

I. Gödel's Theorem

"Any formal system capable of representing arithmetic on the natural numbers, is either inconsistent or incomplete."

Technical terminology:

- *Formal system:* Contains a system of symbols and formal (syntactic) rules for manipulating them. Intended to represent reasoning.
- *Theorem:* A sequence of symbols that can be derived according to the rules.
- *Inconsistent:* An inconsistent formal system is one in which
  - Both P and not-P are theorems.
  - Every sentence is a theorem.
- *Interpretation of a formal system:* An assignment of meanings to the symbols in the system.
  - Intended interpretation: The assignment of meanings that was intended by the designer of the system. A system can have more than one intended interpretation. Systems may also have other, nonstandard interpretations possible.
- *Completeness:* A formal system is "complete" if and only if: Every sentence that is true in all intended interpretations, is a theorem.

II. Why Is G's Theorem True?

Broad outline of Gödel's proof:

1) For any system capable of representing arithmetic on the natural numbers, there exists a nonstandard interpretation on which (some of) the sentences represent claims about the system itself and its sentences.
   - Works by assigning a natural number to each sentence of the system.
   - And correlating an arithmetical property with each possible syntactic operation.

2) There will be a sentence of the system that (a) in the *intended* interpretation is just a claim about arithmetical properties of natural numbers, but (b) in the nonstandard interpretation is a claim about the system, equivalent to this:

   (G) Sentence G cannot be produced by the syntactic operations allowed by the rules of the system.

   - Comments: The sentence will actually say that a certain number does not have a certain arithmetical property. The number in question will be the number that corresponds to the sentence itself. And the arithmetical property will be the property corresponding to derivability according to the formal rules of the system.

3) Sentence G is either true or false.
   a) If true: then G is true and (therefore) not a theorem. Thus, the system is incomplete.
   b) If false: then G *can* be produced by the rules of the system, although it is false. So the system enables one to derive *a false statement of arithmetic*. So the system is inconsistent.

4) Thus, it can be shown that any system capable of representing arithmetic must be inconsistent or incomplete.

Comments about this theorem:

- The *hard* part is establishing (2), the existence of the self-referential statement G. This is extremely non-obvious (but nevertheless true).
  - More precisely: The hard part is establishing the general correlation between syntactic properties and arithmetical properties, such that no matter what the rules of the system are, there necessarily exists a sentence that denies that the number corresponding to itself has an arithmetical property, such that the number would have that arithmetical property if and only if the sentence was derivable according to the rules.
  - This takes many pages of mathematical proof, which we can't give here.
- What does the Theorem *not* say?
  - Not a claim about knowability.
  - Not a claim about provability in the ordinary English sense of "proof".
- Why can't you get around Gödel's Theorem?
  - What if you just add statement G to the system as an axiom? Answer: Then there will be another statement, G', corresponding to the new system.
  - What if you add the infinite series of statements, G, G', G", etc.? Answer: then there will be a new statement, G^{ω}, corresponding to the new system.
- Important note about how the theorem works:
  - Godel does not merely show that sentence G must exist.
  - He shows us *a method of constructing* the sentence G, given any formal system. A mathematician who understood Godel's procedure could repeat it for any formal system.

III. Lucas' Argument

1. If determinism is true, then all human thought could be correctly represented by a formal system.
2. Not all human thought can be correctly represented by a formal system, for:
   a. Suppose that formal system F is alleged to represent all human thought.
   b. F should be incomplete (rather than inconsistent). Reason: if inconsistent, then every sentence is a theorem.
   c. So F cannot derive its own Godel sentence, G. (From b and Godel's Theorem.)
   d. A human mathematician could derive sentence G, using Godel's procedure. (Premise.)
   e. Thus, the human mathematician's thought is not fully represented by F. (From c, d.)
   f. Since this argument could be given for any F, it follows that no formal system fully represents human thought.
3. Therefore, determinism is false.

IV. Objections to Lucas' Argument

Objection #1: Maybe humans are modeled by an inconsistent formal system.

Objection #2: Lucas' argument only shows that we couldn't *know which* formal system represents us.

Objection #3: Premise (1) is false, because

a) Formal systems have countably many (ω) sentences.

b) Physical objects have continuum many states.

c) Godel's theorem can't be proved for systems with continuum many sentences.

Countable infinity: The infinity of the natural numbers. A countably infinite collection of objects can be put in one-to-one correspondence with the natural numbers.

Continuum: The infinity of the real numbers. This is greater.

Phil. 2750
Notes #18: Death

I. Puzzles about the Evil of Death

Issue 1: Can anything be bad other than unpleasant experiences?

Epicurus says:

1. Only pleasure and pain are good/bad for a person.
2. Being dead isn't painful.
3. So being dead isn't bad for a person.

Issue 2: If death is bad, then (a) who is the subject of the harm, and (b) when does he suffer it?

Epicurus again:

1. When you die, you cease to exist.
2. Nonexistent things can't be harmed.
3. So being dead can't harm you.

Issue 3: Why the asymmetry between past & future?

(a) We think our future nonexistence bad. But
(b) we don't think it bad that there was an eternity before we were born when we didn't exist.

Lucretius:

1. Your prenatal nonexistence wasn't bad for you.
2. Past and future are symmetrical.
3. Therefore, your postmortem nonexistence also isn't bad for you.

II. Nagel's View

Why Death is Bad

- Life is intrinsically good.
- Death is bad because it deprives us of life.
  Death is not any sort of *positive* evil.

Many Things Are Bad Other Than Unpleasant Experience

- Being betrayed makes us unhappy because it is bad. It's not bad because it makes us unhappy.
  - Otherwise, there'd be no *reason* to mind being betrayed.
- The case of mental degeneration.
  - Contented infants aren't badly off.
  - This man wasn't badly off when he was 1 year old.
  - Why is he badly off now? Because of what he lost.
- Harms need not consist in present, actual states; they can consist in lost *potential*.

When does the misfortune occur?

- Misfortunes need not have precise spatiotemporal locations.
- They must have subjects, who have precise spatiotemporal locations.

The Temporal Asymmetry

- Time after your death is time you are deprived of.
- Time before your birth is not time you were deprived of.
- [Comment: Compare another temporal asymmetry: We care more about future misfortunes than past.]

Phil. 2750
Notes #19: Death & Aging

I. The Problem of Senescence

- *Senescence:* The natural deterioration of the body due to age.
- What causes it?
  - Telomeres: The physical ends of the chromosomes. Lacking genetic information, but necessary for cell division.
  - During cell division, telomeres are progressively shortened.
  - When telomeres are sufficiently shortened, cell reaches the Hayflick Limit: the limit to the number of divisions possible.
  - Damaged cells can no longer be replaced.
  - Before that point, cells start to lose some of their functionality.
- Some Causes of Death, U.S., 2004:[1]

1. Heart disease	652,000
2. Cancer	553,000
3. Cerebrovascular disease	150,000
4. Lower respiratory disease	122,000
5. Accidents	112,000
6. Diabetes	73,000
7. Influenza + pneumonia	65,000
8. Alzheimer's	60,000
9. Kidney disease	42,000
10. Septicemia	33,000
11. Suicide	32,000
12. Liver disease	27,000
13. Hypertension	23,000
14. Parkinson's	18,000
15. Homicide	17,000
. . .	
AIDS	13,000
. . .	
All causes	2,397,000

- The overwhelming majority of those deaths are due primarily to senescence.

II. The Pro-Death Case

"I also think, perhaps too brazenly, that mortality as such was good for us—I've written on that subject when I was in my 40s—that it's the recognition that we are finite and frail that is the ground of treasuring the things that life has to offer, to make the most of our time, to avoid

[1]U.S. National Center for Health Statistics, <www.cdc.gov/nchs/data/dvs/LCWK9__2004.pdf>, <0-www.cdc.gov.mill1.sjlibrary.org/nchs/fastats/aids-hiv.htm>.

boredom, to appreciate beauty, to spend our time worthily, to love well; that knowing that we don't have infinite time and that we go around only once really is, I think, an absolutely necessary condition for making the most of this unmerited gift of the time that we have. I don't have any great personal desire to live past my allotted time, what it is."

"And so the question would be [. . .] is there some connection between the limits that we face and the desire for greatness that comes from recognition that we are only here for a short time. If you push those limits back, if those limits become out of sight, we are not inclined to build cathedrals or write the B Minor Mass, or write Shakespeare's sonnets and things of that sort."

— Dr. Leon Kass, bioethicist[2]

- Why might death be good?
  - It teaches us to appreciate life more.
  - It makes people do great things.
  - Life would get boring after a while.
  - Mortality is part of our nature. Don't go against nature.
  - Maybe there's an afterlife?

III. Against Aging

- Aging decreases the quality of life.
- It kills you.
- It prevents people from doing more things.

IV. What Could Be Done About Aging?

Could Research Be Successful?

- Many experts think it will soon be possible to rejuvenate middle-aged mice (see Bostrom, note 2).
- We now know essentially what causes aging.
- What might stop it?
  - Telomerase: An enzyme that adds to the telomeres. Believed by some to be the key to an effective anti-aging treatment.
  - Problem: It is also present in 90% of cancer cells, and apparently responsible for their ability to grow tumors.

Some Spending Statistics for U.S., in billions of dollars:

AIDS-related federal expenditures (2005)[3]	20	(3 for research)

---

[2]Kass was the chair (2002-2005) of the President's Council on Bioethics. Quoted from interviews at <www.pbs.org/wgbh/pages/frontline/livingold/interviews/kass.html>, <www.fightaging.org/archives/000084.php>.

[3]http://www.kff.org/hivaids/upload/Fact-Sheet-U-S-Federal-Funding-for-HIV-AIDS-The-FY-2006-Budget-Request.pdf

National Cancer Institute budget (2007)[4]	4.7
National Institute on Aging budget (2007)[5]	1
Total health care expenditures (2005)[6]	2,000
Total GDP (2006)[7]	13,000

V. Why Not More Effort to Stop Aging?

- Aging accepted as part of the human condition. "Normal" conditions don't upset people.
- The dragon viewed as invincible. New technological possibilities not fully appreciated. Failure of previous efforts discourages people.
- Priorities are backwards. Other issues take more attention:
  - More dramatic, unusual causes of death take precedence (terrorism, war, homicide)
  - More *ideologically* charged diseases get priority (AIDS, breast cancer)
  - This suggests that social activists are not attuned to the actual welfare of humanity.
- Rationalizations for appreciating death.
  - Religious views about afterlife.
- Failure to appreciate the urgency:
  "It matters if we get the cure in 25 years rather than in 24 years: a population greater than that of Canada would die as a result." (Bostrom, 12)

[4]http://plan.cancer.gov/

[5]The division of the National Institutes of Health devoted to aging research. Source: http://www.nia.nih.gov/NR/rdonlyres/169BAC4D-FDF9-4FB4-8CB9-1D323F0F83F9/4204/CJFINAL13106PDF.pdf

[6]http://www.nchc.org/facts/cost.shtml

[7]http://siteresources.worldbank.org/DATASTATISTICS/Resources/GDP.pdf

Phil. 2750
Notes #20: Hedonism

I. Hedonism

- *Ethical Hedonism:* The view that pleasure is the sole intrinsic value.
  Intrinsic value: Thing that is desirable for its own sake.
  Instrumental value: Thing that is desirable for the sake of something else.
- Not to be confused with *psychological hedonism:* The view that people are motivated only by pleasure.

In defense of hedonism:

- Note that 'pleasure' or 'enjoyment' may be either emotional or sensory.
- Try to think of anything else that is good. You will probably find that it is only instrumentally good. You think other things are good because they give you pleasure.

II. The Experience Machine (Nozick)

- A machine that can produce any series of experiences you want, by direct brain stimulation. Can be programed with a variety of happy experiences. Can also erase your memory of life before the machine.
- Should you plug in?
  - Most people say 'no.'
  - But hedonism implies 'yes.' Hedonism implies this would be the best possible life.
  - This seems to show that hedonism is false; something matters in life other than pleasure, and other than one's subjective experiences. What? Nozick says: "perhaps what we desire is to live . . . ourselves, in contact with reality."
- Notice how this argument works:
  - Thesis to be criticized: "pleasure is the only intrinsic good."
  - Strategy (sometimes called "*the method of isolation*"): to 'isolate pleasure' (imagine a life with pleasure & nothing else commonly considered valuable). Ask whether it would be good. If not, then pleasure isn't the sole intrinsic good. (Can be applied to anything else that is claimed to be the sole intrinsic good.)

1. If pleasure is the sole intrinsic value, then a life containing a lot of pleasure, without any of the other things in a normal life, would be a great life.
2. But (intuitively) such a life would not be great.
3. So pleasure is not the sole intrinsic value.

Phil. 2750
Notes #21: The Problem of Evil

I. Basic Idea

- The 'problem of evil' is a problem for theists: it is the problem of explaining how an all-good God could allow evil to exist.
- Atheists argue that the existence of evil is a reason not to believe in God.

II. What Is Evil?

- In this argument, "evil" is anything that is bad, i.e., anything that it would be better if it didn't exist.
- Two kinds of evils:
  i) Man-made evils: war, murder, slavery, etc.
  ii) Natural evils: disease, tornados, some famines, etc.
- Some striking examples of evil:
  i) Adolf Hitler, Josef Stalin, and Mao Tse Tung: These three individuals collectively killed probably over 85 million people (!) by means of executions, extreme conditions in slave labor camps, and man-made famines. (Source: <http://www.gmu.edu/departments/economics/bcaplan/museum/faqframe.htm>)
  ii) Animal suffering. Richard Dawkins writes:
  "The total amount of suffering per year in the natural world is beyond all decent contemplation. During the minute it takes me to compose this sentence, thousands of animals are being eaten alive; others are running for their lives, whimpering with fear; others are being slowly devoured from within by rasping parasites; thousands of all kinds are dying of starvation, thirst and disease. It must be so. If there is ever a time of plenty, this very fact will automatically lead to an increase in population until the natural state of starvation and misery is restored." (Dawkins, *River out of Eden*, 131-2)

III. The Argument from Evil

1. There is evil.
2. If there were a God, there would be no evil. For:
   a. God would be aware of any evil. (For he is all-knowing.)
   b. God would be able to eliminate it. (For he is all-powerful.)
   c. God would be willing to eliminate it. (For he is all-good.)
   d. If someone is aware of an evil, and he is both willing and able to eliminate it, then he eliminates it.
   e. Hence, God would eliminate (or prevent) all evil.
3. Therefore, there is no God. (from 1, 2)

Alternate version:

1. If God created a world, he would create only the best of all possible worlds. For:
   a) God would be all-knowing, all-powerful, and morally perfect.
   b) A morally perfect being chooses the best option available to him that he knows about.
   c) An all-powerful being has all possible worlds available.
   d) An all-knowing being knows about all the options.

2. This is not the best of all possible worlds.
3. Therefore, God did not create this world.

IV. Responses to the Argument from Evil

#1 and 2 are common among theologians. #3-8 are common among students.

1. *The Free Will defense:* Evil is a product of human free will. God gave us free will because free will is a very valuable thing. But he cannot both give us free will and prevent us from doing evil.
2. *The virtuous act response:* Some amount of suffering is necessary in order for humans to develop important moral virtues. Some moral virtues can only exist in response to suffering or other bad things. Examples: courage, charity, strength of will.

Reply to responses 1 & 2: Is this a good reason for not stopping evil? The murderer example. These are common among students.

3. *Evil is necessary for good:* Good and evil exist only as contrasts to each other. Therefore, if evil were eliminated, good would automatically be eliminated as well.
4. *Evil is necessary in order for us to understand good:* Slightly different from #3: If all evil were eliminated, then we wouldn't *know* that everything was good, because we can only perceive things when there is a contrast.
5. *How do we know what good and evil are?* Perhaps God has a different conception of evil from ours. Thus, maybe he thinks the Holocaust was actually a good thing.
6. *The Lord works in mysterious ways.* Perhaps there is some underlying purpose served by all the evil in the world, but we humans are not smart enough to comprehend it. Just have faith.
7. *God isn't responsible for the evil. The Devil is.*
8. *Perhaps God is not all-knowing, all-powerful, and/or all-good.* If we simply weaken the definition of God, then the existence of God may be compatible with the existence of evil. Thus, for example, he might be unable to instantly eliminate all the evil.

Phil. 2750
Notes #22: Central Planning

I. Central Planning in Reality

- Central planning: A system in which a central authority (esp., the government) directs all or most economic activities. The authority tells people what to produce and how much. Ex.: Soviet Union, Communist China, Nazi Germany.
- More generally: Social engineers try to predict and control the direction of society.
- Problems in reality:
  - The Problem of Incentives:
    a) Planners may have self-interested, emotional, and other non-altruistic motives.
    b) Producers and consumers lack motives for efficiency, productiveness.
  - The Problem of Knowledge: Planners lack knowledge of:
    a) People's utilities.
    b) Details of individual businesses and products.
  - The Problem of Ability:
    a) No reliable mechanism to select competent planners, remove incompetent.
    b) Market mechanisms:
    Competition: Many different firms in any given area. Stronger firms drive out weaker.
    Limited funding: Firms can only continue their activities if investors are willing to give money. Sufficiently serious failure removes the company from the market through bankruptcy.
    Personal responsibility: Those who create a plan lose *their own* money when it fails.

II. Central Planning in Asimov's Universe

- No problem of knowledge. Psychohistory enables extremely accurate predictions of the future.
  - The analogy of thermodynamics: Behavior of individual particles is unknown. But behavior of large groups is predictable.
  - Similarly, behavior of individual humans is unpredictable. But behavior of large groups is predictable.
  - Except for occasional, exceptional individuals messing things up. ("The Mule")
- The future is also, however, easily manipulable by small interventions.
- The Foundation are benevolent planners.
- Is this a believable story?
  - That the course of history is predictable? Human history may be *chaotic*.
  - That, at the same time, it would be changeable through small interventions?
  - That humans would possess the requisite knowledge?
  - That benevolent planners would exist?

Phil. 2750
Notes #23: Anarchy

I. 'Government' & 'Anarchy'

Important concepts: (Friedman's def's)

- *Coercion:* Violation of what people generally regard as the rights of individuals with respect to other individuals.
- *Legitimized:* Generally accepted as legitimate in a given society.
- *Government*: An agency of legitimized coercion. Examples:

Action	What it's called if you do it	What it's called if the state does it
Seizing someone's property without their permission	Robbery / extortion	Taxation
Forcing people to work for you	Kidnaping / slavery	Conscription, jury duty, national service
Killing lots of people (For a political goal)	Mass murder (Terrorism)	War

- *Anarcho-capitalism:* Society with no government, but with private property; provides 'governmental' services by alternative institutions

II. Police, Courts, & Laws under Anarchy

- How would these presently governmental institutions be replaced?
- 'Police':
  Private security guard companies protect people from criminals. There are multiple competing companies in the same area.
- Courts:
  Private arbitration firms are used to resolve disputes. There are multiple competing arbitration companies.
  Private contracts specify arbitration agreements.
  Protection agencies sign arbitration agreements with each other.
- Law:
  Laws are made by judges/arbitrators. *Note*: compare the British common law.

III. Advantages of Anarchy

A. Non-coercive:

Governmental system: you are forced to accept a government, and have little control over what kind of government you have.

Anarcho-capitalism: You choose whether to hire a protection agency, and which one.

B. Less risk of abuse of power:

Governmental system: Government has a monopoly → There is no one to stop the government from abusing its powers.

Anarcho-capitalism: Competition among many protection agencies → Customers can leave a bad agency.

C. More efficient:

Governmental system: Police have no incentive to reduce crime. If crime rates go up, they get more money & more police are hired.

Anarcho-capitalism: Protection agencies have an incentive to reduce crime. If crime goes up, they may be fired.

Empirical observation: The free market is more efficient at producing food, shoes, computers, automobiles, etc., than the government is.

D. Removes public goods problem:

Governmental system: Informed voting is a public good. Laws & protection are public goods.

Anarcho-capitalism: Good laws & protection are private goods.

IV. The Problem of Monopolies

• *"Natural monopoly":*

Occurs when the optimum size for a firm is so large that there is room for only one such firm on the market.

This situation is very rare.

Even natural monopolies are restrained by "potential competition."

All products compete with all other products.

• *Artificial monopoly: the strategy of 'predatory pricing':*

Problems:

Larger firm loses more total money.

Larger firm may be less efficient.

The monopolist must sell to everyone, at a loss; small firm need not sell anything at a loss.

Also forced to increase production, losing even more money.

• *Cartels:*

Have all the problems of monopolies.

Additional problem: 'chiseling'

Compare: why don't all the farmers refuse to feed anyone unless everyone agrees to give them everything?

• *State Monopoly:*

Almost all actual monopolies are government-enforced. Government agencies get taken over by the industry & used to the industry's advantage. Why:

They are the ones who know most about the industry.

They have the most incentive to try to influence the agency. Influencing policy is costly, time-consuming.

Examples:

The Civil Aeronautics Board

The American Medical Association

Could this problem be solved in a better government?

The pattern is not an accident. It is built into the logic of the system.

V. Further Questions

1. Wouldn't the security agencies fight with each other?

- They would take disputes to the arbitration firms, because this is economically rational.
- *Contrast:* what happens when governments decide to fight each other?

2. Why would agencies obey an arbitrator's decision?

- Companies abide by arbitrators' decisions; otherwise their reputations would be ruined.
- Violating decision defeats the point of going to arbitration.

3. How could you know the arbitrator was fair?

- There is competition among arbitration agencies → you can choose a reputable firm.
- *Contrast:* What do you do when the government's courts are unfair, irrational, inefficient, etc.?

4. What if one security agency decides to defend murderers, thieves, etc.?

- Their clients would constantly be costing them money. (Compare: "The Arsonists' Fire Insurance Agency.")
- They fight a constant war against the rest of society.
- They must pay higher wages to their employees.

- The "Thief Protection Agency": They must charge their clients more money than the stolen goods are worth.
- *Contrast:* What happens if you get corrupt people in the government?

5. *Why would security agencies protect the poor?*
   - Why would this be different from any other good or service? Why do food companies feed the poor?
   - They're already paying for protection (from the state). Private protection would be cheaper and more effective.
   - *Contrast:* Why would the government protect the poor? How well do they in fact protect the poor?

6. *Wouldn't other countries attack the anarchists? Possible answers:*
   - Ideally, the whole world would be anarchist.
   - Almost all wars are due to (i) disagreements between governments, (ii) racial and/or religious hatred, and/or (iii) perceived historical injustices. (Compare: Why don't other countries attack Switzerland?)
   - It is harder to take over a territory with no government, than one with a government. Attacking countries use the governmental structure already in place to control the populace.
   - The security agencies & the general population would be armed. (Compare: U.S. experience in Vietnam.)