Notes on Death

Sam Enright

(Adapted from PHIL 176 with Shelly Kagan)

Notes on Death .......................................................... 2
Arguments for the existence of the soul .................................. 2
Plato’s Phaedo ...................................................................... 2
The nature of personal identity ............................................. 4
Belief in death .................................................................. 6
Is death bad? .................................................................... 6
Immortality ....................................................................... 8
The value of life .................................................................. 8
Fear of death ..................................................................... 9
Suicide ............................................................................ 10
Writing Advice .................................................................. 11
Notes from the Readings .................................................... 11
Review of the Course ........................................................ 16
Notes on Death

Arguments for the existence of the soul
The most popular arguments for the existence of the soul might be characterised as such:

- There must exist a CPU; a controller to the body and our actions. Machines have controls that are external to their physical forms, and humans have internal ones.
- The existence of free will.
- The existence of reasoning. This argument has generally been retired now but people make the same argument with feelings, or sometimes creativity.
- Descartes argued for the existence of a soul, or something similar to it, by arguing that the mind and the body were logically distinct. You can tell a story – conceive of – a situation in which your soul exists, but your body does not, but you can’t tell a story in which a chair exists and also doesn’t exist. Hence your body and soul must be logically distinct (even if they are not physically distinct). There is still active debate as to whether any form of this argument is valid. A similar argument comes up in the discussion of philosophical zombies, namely, if we can conceive of a zombie functionally equivalent to a human being with no subjective state whatsoever, what is the significance of this? What exactly do we mean by conceivability? Depending on how you read the argument, it could be the case that it’s merely possible that the mind and brain are separate, but they are in fact the same in this world. Dualism would thus be contingent on what happens to be the case in our world.

Plato’s Phaedo
In the Phaedo, Socrates remains unphased by his death and is comforted by his conviction of the immortality of the soul. Over the course of the dialogue, he offers up the following arguments for this:

1. The argument from the nature of the forms: The forms are not physical objects. The only thing that can grasp them must be nonphysical, in this case the soul. While this is a dubious argument for the immortality of the soul, it might lead you to wonder how the death of the body could ever cause the death of the soul (since it is nonphysical). However, this resolution to the objection is weak, because sensory organs cause things in the soul all the time (if not perceptions, then at least changes in character). We can add an additional premise to this argument; namely that forms are eternal, and thus can only be grasped by that which is eternal, in this case the soul.

   The most objectionable step in this argument is positing that something must be eternal and nonphysical to grasp something that is eternal and nonphysical. A biologist can study a cat without herself being a cat. This is an imperfect objection, however, because cats and biologists are in the same category of thing. I think the more powerful objection comes from how the soul can grasp the physical world but not the other way around. If the supposedly eternal and nonphysical soul can understand the nonphysical world, what makes us think that nonphysical eternal things can be contemplated by nothing other than themselves?

2. The argument from ‘recycling’: things in the world are made out of things that previously existed. When you die, your soul remains as an indivisible part. The soul only changes (just
as it goes from the realm of the awake to the realm of dreams and back again, an example Plato gives). This is pretty similar to the atomic idea put forward but Democritus.

The objection here is that the soul may not be indivisible. We would not say, for instance, that the heart is immortal, because after you die your body decays and the material from it goes into other things. While it’s true that there are smaller components of the heart that would remain intact (cells, tissues, atoms), the heart itself would cease to exist.

3. **The argument from recollection:** just as a photograph can remind us of an individual, Plato’s theory is that physical things in the world embodying the forms to a greater or lesser extent remind us of those forms. But, for something to remind us of something else, we must already be acquainted with it. Thus, we must already have had an acquaintance with justice, beauty, and the other forms before our birth. Thus, if the soul survived to enter a new human body, there is no reason it must die with it.

An alternative explanation is that we are born with an innate knowledge of forms, similar to the innate morality we are born with. However, since there are infinitely many possible forms (I assume?), it’s hard to see how these would all be manifest in a finite brain. This is one strength of the ‘acquired-before-birth’ theory of the forms: that it could involve finitely many encounters with forms that manifest themselves in your brain. This skirts the broader issue of how a soul could have memory. What would count as a mental representation within a soul, and if it immaterial, how are memories encoded on it?

4. **The argument from simplicity.** In the *Phaedo*, there is an unstated worry about how, even if we accept the ‘recycling’ argument, and grant that the soul was around before our birth, that does not give us reason to believe the soul can’t break. Although this is unstated, Plato seems aware of it, as he offers up this argument. Plato turns to a discussion about what can and cannot be destroyed. The things that can be destroyed are those who are constructed out of simpler parts. The forms therefore cannot be destroyed. We know that a form is not composite because Plato asserts that only changing things can be composite, and the forms are unchanging (the number 3 will always be odd...). Since invisible things don’t change (?), the soul cannot be destroyed. Socrates doesn’t actually conclude this; he concludes that it is invincible... “or nearly so”. Simmias offers the counter-example of harmony as an invisible thing which can be destroyed. Socrates then spends some time criticizing this analogy. This is a flawed strategy given that so long as harmony is an invisible thing which can be destroyed, one of his premises is incorrect. One comment about this observation is that it is, in a way, a statement of the essence of physicalism; that the soul is something that can be created, destroyed, and emerges from the interaction of parts.

There are a few things that Plato might refer to when he says that something is invisible; that it can’t be (a) seen (b) observed (c) detected. He can’t quite mean it in sense (a), because harmony can’t be seen but can be destroyed. There’s good reason to think that he meant it in sense (b). The sense (c) might be needed to modernize the argument, now that we know there are things like radio waves which are destructible and physical in all the usual senses, but outside the scope of our senses. However, broadening to case (c) creates a problem: is the soul still invisible? If it is, how could you ever know that you had one?

Within the scope of Plato’s works, the argument for the simplicity of the soul is strange. In *The Republic*, he famously argued that the soul has at least three parts: the rational, spiritual (as in, the will), and the appetitive.

Plato responds to the harmony objection by noting that harmony does not exist prior to the physical construction of an instrument. But, as he has already argued (the
argument from recollection), the soul exists prior to the physical construction of the body. An instrument can also be more or less harmonious, but a soul cannot be more or less soul-y. However, if we conceptualize the soul as something more like consciousness or the intellect, it can; in partially wakeful states, for instance. You can also have a harmony of harmonies, but not a soul of souls. Also, the soul can affect the body (and vice versa) but a harmony has no causal power over the notes that comprise it.

5. **The argument from essential properties:** contingent properties are those which are inessential to an object being what it is, and essential properties are the opposite. A red car could have been blue, but a fire could not have been cold. Plato says the essential property of a soul is being alive, or to cognate. The obvious objection here is that heat being an essential property of fire says nothing about how long the fire must go on for. But this is a special case: since the soul has an essential property of life, it is deathless, so it must be immortal.

There are two possible conclusions here: either it can’t be the case that the soul exists and is dead, or is can’t be the case that the soul was destroyed. The latter case is much stronger; it forbids the case of a fire being extinguished even though it must be hot to be a fire. In Kagan’s view, Plato somewhat flip-flops between these views.

Socrates actually thought that bodily sensations were taken care of by the body, unlike the modern dualists. He believed that the soul was what thinks. The soul is what’s capable of conceptualizing forms. How can we think about beauty itself? Plato would say that individual things may participate in beauty, but given that none of them are perfectly beautiful, how do we know what beauty is? A similar argument can be applied to numbers: how can we really know what two is, given that we have only interacted with sets that participate in two-ness? Many philosophers still think that the basic insight behind Plato’s metaphysics – namely, the idea of forms – is correct.

Plato thinks that the empirical world has somewhat the same character of insanity as dreams. Things are big (a basketball player), while being small (only 6 foot), hot while being cold, etc. The world is constantly rolling between one form (in this case, bigness) with another (smallness).

A soul can be better or worse at getting closer to the forms. The body and primal urges are concerned primarily with the empirical world, but by training ourselves to be better at appreciating forms, Plato argues there is a greater chance that the soul will escape the body after death. A question that one might ask about Plato’s metaphysics: is it true that you have to have contact with the forms at all to think about something? Do I need to have had some acquaintance with perfect straightness to think about a straight line? Or can I simply extrapolate from cases in which I know a line is more or less straight? In modern parlance, what cognitive mechanism would explain an inability to do that? Even if we grant this point for the time being (forms are necessary to think about a subject), we have a second question: why can’t I acquire knowledge of a form in this life?

### The nature of personal identity

There are various theories as to what causes someone to have a constant personal identity:

1. **The soul theory:** someone remains the same person as long as they have the same soul. Among other issues, this raises the problem of what would happen if your soul were suddenly replaced with one which was indistinguishable. You don’t know whether this has happened to you, creating problems for a notion of personal identity at all. Locke raised this
issue, and thought it was too large a pill to swallow; that there would be no way of telling that your soul had been replaced.

2. **Body persisting through time:** An alternative theory is that personal identity is tied to your body persisting through time. This leaves the only option for life after death as being bodily reconstruction. Parenthetically, there have been cultures that believed in this notion of personal identity, in particular in the early Christians, who thought God reformed your body on judgement day. Peter Van Inwagan has a thought experiment of knocking over your child’s tower of blocks before your wife comes home, then rebuilding it in exactly the same way, and showing it to her. Would it still be the same tower your child built? The intuition that it wouldn’t pushes back against the body theory. Note that we could solve this by adding a condition of spatiotemporal contiguity to our bodies.

3. **Brain persisting through time:** The most popular and coherent version of the body theory is that personal identity consists of the brain being the same, but not all body theorists believe this. But this is, in practice, problematic; since there are many disorders and accidents which cause a large chunk of the brain to become non-functional, or to be repurposed, as in the split-brain experiments.

4. **Personality persisting through time:** If we recall the Lockean worry about how arbitrary it was that our identities could change without any of our characteristics changing if our soul were replaced, we might be tempted to just say that personal identity is a matter of personality – an amorphous set of desires, intentions, memories, and temperaments – persisting through time. It’s important to note that rejection of the soul is not synonymous with the rejection of the soul as the root of personal identity. Locke, for instance, thought that we had souls, but that our identity didn’t depend on them.

A thought experiment we might run to determine where exactly our intuitions lie would be to separate out your body and personality, and ask who (or which) would you rather be tortured? Would you feel any (or much) better about being tortured if an evil scientist completely wiped your memory beforehand, and thus destroyed (at least part of) your personhood in the personality view? We might also add an addendum to the personality view, namely that you would have to, in fact, have experienced the things you thought you had experienced to qualify for being the same person, which sidesteps problematic cases like insane people who believe they are, say, Napoleon. If we don’t want to go this far, we could add a ‘no branching’ caveat, namely that someone’s personality cannot branch into multiple versions. Since we can’t say that any one of these is the individual, we can’t say that any of them are. There is trouble with cases where one half of your brain goes into one body and one half into another. This might require us to add the ‘no branching’ caveat to the body view also. In the philosophical literature, branching is referred to as fission.

The soul theory has just as much trouble incorporating the ‘no branching’ rule. It may in fact have more difficulty because, if your soul went into just one body, there would be no way to tell which one, either from the inside or the outside. This raises epistemic difficulties (if there is no way to know something even if principle, is it really knowledge?) that would not arise in the case of, say, identifying which cognitive functions went to which body in the split-brain case. Recall that the motivation for positing a soul is that you need souls to explain people. But this means that, if a person splits in two, then their soul splits in two.

If this all seems a bit abstract, you might refocus the question slightly to what traits or properties matter in the persistence of an identity through time. Even if you think you had
conclusively shown that the body theory was correct, it would seem odd to treat someone with a new personality and total irreversible amnesia with the same brain just as you treated them before. Questions of personal identity might then be relevant to questions about how responsible people are for things they committed a long time ago in their development, or how we would conceive of character and responsibility among beings who lived for many hundreds of years.

Belief in death
People sometimes say that we can’t really believe that we will die, because it is impossible to imagine being dead. Freud considered it a core idea of psychoanalysis that you couldn’t really believe in death. But this is a confused claim: we can’t imagine what it is like to be a chair, but that is because it is not like anything to be a chair, not because we don’t really believe in chairs. If this were true, we wouldn’t really believe that we would ever dreamlessly sleep. You also can picture your death, from the outside.

A common retort to this is that picturing things from the outside is really picture them from the inside: if you imagine your funeral, say, you are implicitly imagining that you exist, because that is who is doing the observing. We must do work therefore to ascertain whether an observer of some mental image is necessarily you.

This tension between explicit and implicit beliefs comes up in domains outside of death. For instance, someone with OCD may wash their hands compulsively while having no explicit belief that their hands are actually dirty. One hypothesis is that they, deep down, do believe that their hands are dirty. People who have brushes with death often act in profoundly different ways afterward, which gives credence to this psychological account of an explicit vs. implicit belief in death.

Is death bad?
Why is death bad? One suggestion is that death can’t be bad for the individual who dies, because nothing is good nor bad for that individual. We’ll call this the Epicurean argument (it was articulated by Lucretius). I think this “worse for no one” reasoning is confused, but I will put that aside for the moment.

If the Epicurean argument is true, the badness of death consists in the effects that that person’s death has on their loved ones and society. One might be used to thinking of bad things as being intrinsically bad, and this is the respect of badness which death falls down on. Death, however, may be a kind of relational badness.

If we assume that there are intrinsic and relational ways in which things can be bad, we can turn to the deprivation account of the badness of death. This says that death is bad because of its opportunity cost. But this brings up issues with the temporality of the badness of death: if death is bad for us, when? Possible responses:

- Death is bad, and it’s bad for us at some specific time.
- Death is bad, and its badness is not tied to any time. Could there be such things, whose truth or attributes are not tied to any time? We shall illustrate one example also relating to death. Suppose I shoot someone on Monday, and I die on Tuesday, and they die from their injuries on Wednesday. When did I kill them? It seems wrong to say that I killed them on Monday, since they didn’t die, but equally wrong to say that I killed them on Wednesday, because I wasn’t even alive. If we clarify ‘kill’ to mean ‘begin the causal path of events that leads to the death of an individual’, this is not helpful, because there are many different points at which we may say that a causal path began, and this is dependent on our views on free will. So,
going beyond semantics, it seems that there is no fact of the matter as to when I killed this person.

- Death is bad, and it’s bad for you when you’re dead. In everyday speech, we describe things being bad for us even though we don’t know about them all the time, which is functionally equivalent to not existing. For instance, if your spouse cheated on you and you never found out about it, most people would say that that was bad for you (in a sort of relational way), despite you being totally unaffected by it. This might be the case for death also. The issue with saying that things can be bad for you even if you don’t exist is that it gives us moral concern for all of the merely possible people, of which there is an enormously large amount. Saying that something can be bad for you if you don’t exist but only if you’ve never existed is some questionable layering of metaphysical assumptions.

We may state the Epicurean point another way: why do we care so much more about loss of life that we had than loss of a possible future of life (what Kagan calls “schmoss”)? The psychological component of this is interesting: why do we care so much more about losing things than about failing to get them, even when the situations are mathematically identical?

We can solve the loss-schmoss asymmetry by appealing to personal identity. For, if you were born any sooner, you would not be the same person (different sperm, etc.), but if you had lived longer, you would have lived longer as the same person. I don’t think this is a complete resolution because many people mistakenly believe that they would be the same person if they were born earlier, and yet these people are influenced by loss and not schmoss. Nagel points out that you actually can imagine someone being born earlier than they actually are, namely as a result of the existence of sperm and egg banks. Ought people born from thawed eggs and sperm lament that they were not born earlier?

Feldman points out that when we think of the possibility of being born earlier, we are not really thinking about the possibility of living longer, only the possibility of shifting our life so that it begins sooner. And at this level, a combination of status quo bias and an implicit understanding of human progress (despite the “born in the wrong generation” rhetoric) may mean that we are more or less all happy to have been born at the time we were. But suppose that instead of shifting life back in time, we imagine that an asteroid is going to hit the Earth next week. We may then feel regret that we were not born earlier, and symmetry is restored.

Parfit comments on this issue, pointing out that the loss-schmoss asymmetry is just a subset of the more general psychological tendency to care about the future much more than you care about the past. He gives the following example: you are told that you will be given an extremely painful operation without anaesthetic, but then given a pill to give you localised amnesia, so you won’t remember it. Now imagine you wake up in the hospital and ask the nurse whether or not you have already had the operation. The sheer magnitude by which you would prefer to already have had the operation than to be scheduled to have it is enormous. Parfit wants to have already had the operation. Daniel Kahneman has discussed some of these themes, in particular his TED talk ‘the riddle of experience vs. memory’, in which he offers the thought experiment of whether you should choose to have a just ok holiday, or your dream holiday where at the end you take a pill that will make you forget all of it. Since the hedonic value gained from memories is small (certainly not enough to bring an ok holiday to a dream one), our decision rests on whether we view happiness – or whatever we are striving to maximise in life – as the sum of momentary pleasures or more broadly as a kind of sense of meaning. This is beyond the scope of this course, but suffice to say there is a large psychological preference in favour of the future rather than the past. Even so, just knowing that this is indeed an attitude does not mean that it is a rational attitude.
We might ask the auxiliary question of whether the inevitability of death makes it better, worse, or keeps it the same. Some people might say that there is no point crying over spilt milk, and so the power death has to upset us is lessened by our non-control over it.

People sometimes make the claim that everyone dies alone. On a naive reading, this is ridiculous. But one way we might make sense of it is to say that no-one can take your part in your death, while they can in other activities that you do co-operatively with others – like playing music. But on further reading, this reveals itself to be a trick of language. No-one can die your death, but similarly, no-one can write in your handwriting, since it is by definition produced by you.

**Immortality**

“I don’t want to achieve immortality through my work. I want to achieve it through not dying”  
*Woody Allen*

Note that even if we completely buy the deprivation account, it does not logically follow that immortality would be a good thing. We would point out that, for any good thing, there is an amount which is too much. The same might be true of life.

I’m not sure what the deprivation account says when a life is *really* bad. Does it say that death is bad because it deprives us of any good things, or because it probably deprives us of a life that is mostly good? I think this is relevant to the immortality case. Living forever is a psychological experiment that’s never been run before, it might be that centuries of grief and trauma would drive any person to insanity, and that, in the absence of reliable drugs to induce amnesia, the immortal person descends into unimaginable levels of torture. The question also depends on our conception of immortality: it would be torture to be immortal if the body still aged and gave you the non-fatal health difficulties of old age. This is a point made by one of the imagined lands in *Gulliver’s Travels*, and by Montaigne. Similarly, eventually civilisation would collapse and the universe would come to a heat death, which is a challenge to the very concept of immortality. But if we put aside all of this, we might look at the in principle rather than in practice question of whether immortality could be a good thing.

Immortality plays an interesting role in global culture and mythology, in that it seems great at first but not so much when you try to fill in the details. It’s not surprise then that religions are so vague on the details about the everlasting life that is promised by many of them. Indeed, it is good that they are light on details, because, when you think about it, is there really any life that it would be good to have forever? Bernard Williams, in an essay on death, argues that no such life fulfils this.

Even if we think that no immortal life would be a good one, there is presumably still a best immortal life. This would probably require large psychological transformation so that you wouldn’t get bored or feel a sense of meaninglessness – maybe involving amnesiac drugs.

So, if we are convinced that immortality would be a bad thing, then it is good that we die. So, death is bad when we die but it’s good that we die. The best case may be a kind of immortality where you can choose to end your life at whatever time you wish. It’s certainly my intuition that most of us die far too soon.

**The value of life**

Let me start with a distinction. Things can be *intrinsically good*, or they can be *instrumentally good*, that is, good because of the things that they lead to. Hedonism says that pleasure and pain are the *only* intrinsic goods. A supposed refutation of this is the Robert Nozick’s experience machine.
The hedonically-minded talk sometimes about a net-positive or a net-negative life (such a concept makes a lot more sense when all that really matters is pleasure and pain). Though I have some leanings this way, I would like to be able to speak about ethics without making reference to a net-negative or a net-positive life. Psychologically, this may be why the Repugnant Conclusion seems so, well, repugnant. It’s extremely hard to imagine someone having a life which is just the tiniest bit better than neutral, and to understand that that life would indeed be positive.

Some people would say that there is value to life irrespective of what that life is like. It’s hard to see how this would still be true of people in persistent vegetative states. Note, however, that this does not mean that life is always worth living; there might be a finite value to life which is comparable with pleasures and pains. One could make the stronger claim that life itself is infinitely valuable, but I think this is philosophically incoherent on a number of fronts. We might also say that the overall narrative arc of a life matters to the quality of that life. We might say that a rags-to-riches life is better than a richest-to-rag life, even if they last just as long and are hedonically indistinguishable.

It is a puzzle to explain why people do the risky things that they do. The views from skydiving for instance, could be achieved almost as well just looking down from an airplane, and span such a short period of time that you would have to put a very low value on your life to think the risk was worth it. Yet people who value their lives very highly still subject themselves to this risk. Parenthetically, I find the subject of risk in psychology and economics fascinating, and have more thoughts I want to develop on this subject.

In questioning whether it’s worth it to be born, we’re asking whether the goodness of life exceeds the badness of death. But it’s not clear that we can just add the effects of life and death together; there may be interaction effects. Some plausible interaction effects include that it is bad to get a taste of life while having such limited experience of it and that it is bad that we decline so much from being alive to being dead.

The outlook that life is good and would be bad to lose can crudely be described as the Western approach. The idea that life is not as good as we think it is more expressed in the Eastern tradition (such as Buddhism, with its mantra ‘life is suffering’).

Fear of death
Is it rational to fear death? Well, when we describe an emotion as being rational, it is usually because it meets some conditions. It makes sense – in the logic of emotions – to feel pride about your country doing well at the Olympics, but it doesn’t make sense to feel pride about a country you have no connection to doing well at the Olympics¹. It also makes sense to feel pride about something which you are responsible for.

Kagan argues that it does not make sense to be fearful of the inevitable. You could rationally be sad or resentful about it, but not fearful. Personally, I find this formation of necessary-and-sufficient conditions for emotions to be unhelpful. Death either does or does not lead to an emotion that we would call fear. Understand that I am not bringing into question the ontological status of an emotion like fear, rather questioning whether the particularities of the English language reliably map on to a specific set of psychological symptoms, and whether this set is the appropriate one to be talking about. We can talk about ways to prevent or provoke this, but I don’t think the question of whether it is rational to fear anything makes sense. Nonetheless, I’ll bite: fear comes from

¹ This is a sense of the term pride that has troubled me in the past. The closest thing I can come to a resolution is to say that the responsibility variety of pride is different to national or familial pride. The latter expresses more of a gladness and an approval that this is a group of which you are a member.
uncertainty. Given that we know that we will die, it does not make sense to be fearful about it. But we don’t know when we will die or how we will die, so even this formulation is problematic.

Epicurus famously made the argument that the reason we seek out material success is because we want people to remember us, and thus because of an implicit fear of death. The fear of death, among other things, might lead one to argue that talking and thinking about death is positively harmful. Knowing certain facts about death may cause us to change our behaviour but note that this information may or may not change our behaviour for good reasons. Information about the digestive and excretive processes of a partner may make you not want to kiss them despite giving you no new information or reason to not do so. If information about death is of the latter category, we might argue that it would be better off if we never talked about death. It’s an empirical question whether atheists or philosophers adjust to their own death better than people who have false beliefs about death or don’t think about it.

Suicide

How could you say that you would be better off dead? Even if we agreed that it would better off never to have been born (as anti-natalists believe), this is not an argument for suicidality. The argument from negative interaction effect, namely that we get such a brief taste of life that it would be better off never to get, doesn’t condone suicidality because suicide would make it an even shorter taste. There are really two questions when it comes to the rationality of suicide: whether it would ever make sense to kill yourself and whether you would ever be of the right frame of mind to make that judgement. But, in fact, we accept the judgements of others as valid even when they are in a situation that clouds their judgement, like someone making the decision to have a surgery that carries a risk of death.

Some people make the theological point that it is against God’s wishes to commit suicide. But if the issue is interfering with God’s intentions for who lives and who dies, then there is similar moral issue with saving someone’s life. Another argument says that life is a gift, and to throw away that gift shows ingratitude, which is wrong.

A deontologist might also say that suicide is forbidden because it involves killing. But we might respond to this by saying that morality is about how you treat others, not how you treat yourself. But why would we as individuals be exempt from our own moral thinking? It’s possible that the deontological ban on harming an innocent is really a ban on leaving someone worse off than they were before (few people would be willing to say that a surgeon acted wrongly if he amputated your leg to save you, for instance). In this case, assuming this suicide is a rational one, you are actually made better off by being dead. If the deontological dictum to not kill extends to yourself, it would also forbid cases of self-sacrifice that we would regard as heroic, such as jumping on a grenade to shield your fellow soldiers.

We might also think about the moral relevance of having the consent of the victim. When you commit suicide, you have your own consent, almost by definition. It’s difficult to put your finger on the exact moral value of consent. If someone were to tell you that you have their permission to kill them, most people would agree you had acted wrongly by doing so – unless you had waited some time, and they were really sure, and even then it’s still controversial. There are many other confounding variables: is the person in the right state of mind to make this judgement? How do they understand consent? It seems that consent is not absolute and has a finite amount of moral weight that can be overpowered.
Writing Advice

You should really be able to give a one-sentence summary of the view which you are defending in any philosophy paper. It might be worth having such a sentence above a piece while you’re writing it.

In general, students present far too many arguments when writing a philosophy paper. It is much better to make one or two truly compelling arguments in detail. It is also best to address the one or two most devastating criticisms. After you’ve finished a readable draft of a paper, show it to someone and ask them to explain to you what your argument is. If they can’t do it, you’ve probably done a poor job.

Kagan says he has never seen a paper which goes too far in making the arguments explicit, but hundreds that don’t go far enough. So frankly, numbering or bulleting premises, or giving them a code to refer back to, could be a good idea.

Notes from the Readings

I was unable to find any readings in red.

• There is some serious clarificatory work that needs to be done to make sense of the question “is there life after death?”. In particular, it requires establishing a notion of life – or of you – that could conceivably persist beyond the physical destruction of your body and even your brain. It thus requires positing something akin to a soul – and the solving of many auxiliary theoretical issues.

How to Write a Philosophy Paper [text]
• See above.


Plato, Phaedo, Indianapolis: Hackett Publishing Co., 1977 (Lecture 6-9)
• See above.
• The Phaedo is nominated by Socrates’ disciple Phaedo, hence the name.
• In the Jowett translation, the introduction mentions the Platonic doctrine of reminiscence, which seems to argue for the existence of the soul based on the fact of latent knowledge, including of mathematics, which can be induced in an unlearned person by showing them a diagram. This knowledge is not quite innate, given that it requires a process of reminiscence, and so it must be given to us before birth.

Perry, John. A Dialogue on Personal Identity and Immortality (Lecture 10-12)
• The characters in this imagined dialogue are Weirob, Cohen, and Miller.
• One of the themes in this essay is the distinction between remembering and seeming to remember. One of the interlocuters is trying to demonstrate the possibility of personal
identity persisting after death. He points to Locke’s argument in An Essay Concerning Human Understanding which states that what relates one person-state to another is the possibility of latter person-states to remember earlier ones. But Weirob, the dying character who is in bed, objects that someone who is hypnotised will only think that they remembered something, while not actually remembering it. Hence Miller’s argument that we can conceive of future people with our identity, and thus can imagine existing in heaven, needs more work.

- Cohen, at this point, tries to rescue Miller’s argument by pointing out something about the hypnotist case: namely saying that real memory = apparent memory + identity begs the question, because identity is the concept in question. We might define ‘real’ memories more narrowly as being apparent memories that have some appropriate causal path between them and the events that they involve (molecules in your brain wiggling around as a result of some event occurring, rather than a hypnotist, for instance).

- Weirob objects that God could make an indefinite number of heavenly beings, all of whom would appear to be her, which would seem to mean that none of them are. This does not rule out the possibility of surviving death, however, because God could make just the right number (namely, one). But this objection does require a revision to the memory theory of personal identity, namely that identity = memory + lack of other remembers. Also, if we find ourselves thinking that we are person X because we remember they were person X too. If we broaden out this objection to the whole universe, you would really need to know the state of every being in the universe and their memories to know that you are you, at which point the concept of personal identity has become useless.

- The group’s discussion turns to the nature of personal identity. The men are arguing for a memory theory of identity, which explains how you can know that you are you without having physical experience of your body.


- This passages poses the question of whether there are ways to destroy an animal or human without killing them. For instance, if a new medical process is performed on a dog to change all of its genes, and then the dog is dismantled atom-by-atom, the dog surely has been destroyed by the time it is a pile of formless cells with a different genetic structure. But there was not point at which it died.


Tolstoy, Leo. The Death of Ivan Ilych (lecture 15-16)

- This novella plays on the idea of death coming as a surprise and unreflective people not really believing that they are going to die. Ilych is shocked by his own mortality.

- This novella was written during a 9-year meltdown Tolstoy had after the publication of Anna Karenina, in which he didn’t publish a word of fiction.


- **Preface:** Nagel thinks that, since the Logical Positivists, there has been a tradition of over-systematising within philosophy. He also thinks that there has been too little reliance on intuition and too much paring down to simplistic theories. Nagel diagnoses the discomfort people feel in suspending judgement. When they are criticised for doing so, they point to how they change their opinions in light of evidence (which they very well might do). This has led to two trends within philosophy: the relentless creation of dichotomies and categories, and analytical theories that try to answer too many questions.

- **Death reading:** You might think that we object to death because of the period of nonexistence that it entails. But we weren’t alive for billions of years without it bothering us. Also, if we were suspended such that we had a discontinuous life that continued in a few hundred years, we would not lament the period of time for which we were not “alive” (although we would lament other things, like the death of loved ones and changes in culture). People who believe that death is intrinsically bad would also have no particular objection to states of unconsciousness, like dreamless sleep, which are subjectively indistinguishable from death. So, if we are to say that death is bad, it would have to be because death deprives us of life. How can we say this? We can separate this into 3 problems:
  1. Can anything be bad for someone if it consists just in depriving her of things?
  2. How do we assign the misfortune of death to a particular individual (see below)?
  3. The aforementioned asymmetry between how we are bothered by the prospect of our nonexistence in the future, but not by our nonexistence in the past.

    - The Epicurean position which is made reference to here – that death is not bad because, for something to be bad, it has to be bad for someone – rests, I think, on a faulty premise. There are many things that we would agree are bad despite being bad for no one. For instance, if a malicious actor increased your risk of cancer, and you subsequently fell ill with cancer, they did not cause you to do so. You could have suffered from it anyway, and there is no fact of the matter as to which individuals had their cancer caused by this malicious actor and which would have had it anyway. There are no identifiable people for whom this is worse, and yet it clearly is! Similarly, we would scoff at a politician who did not care about preventing extinction from nuclear war, because if all humans went extinct, there would be no-one for whom this new condition was worse.

- **A note on point (2):** if we say that assigning harm to an individual is a necessary prerequisite to something being bad, it puts a lot of restrictions on our ability to condemn things like mockery and betrayal if their subject never finds out about them. Objection (2) confines us to saying that the goods and evils that befall people are limited to nonrelational states at a particular instant. But, for instance, it would be bad if your mental state were reduced to that of a child, despite it not being lamentable that you were once a child. Nagel uses this point to argue that there are certain goods and ills that are intrinsically relational.

- **Point (3) is resolved by saying that, at the level of the individual, you could not have been born substantially earlier, but you could have lived longer while still being the same individual. Nozick offers a thought experiment: we discover spores which are what causes humans to be born, and which exist from thousands of years before those humans. If there were a scientific discovery by which the spores could be made to be hatched thousands of years before the normal time of human birth, extending those lives radically at no cost to the
quality of those lives, then one could be said to be actually harmed by not coming into existence earlier.

- Nagel writes about premature death, mentioning how it is a greater shame that Keats died at 24 than that Tolstoy died at 82. But the fact that it is a greater shame is not evidence that it is not also a shame that Tolstoy died at 82. This raises the question of whether something harms a species if it is endemic or natural to it. Is it bad that moles can’t see? Nagel takes the position that the inevitability of something within a species does not make it better. If it was absolutely guaranteed that all lives would end in 6 months of terrible agony, its inevitability would be of little comfort to us. And indeed, if life expectancy were 1000 years, it would be a great tragedy for people to live for only 80 years. Perhaps that same tragedy is occurring all the time, but are better at justifying it.
  - There’s an interesting problem with the concept of a premature death. If we take a premature death to mean one that falls short of the expected years of life someone has left to live, then all deaths are premature. Even if someone is past the life expectancy, there will be an expected number of years they have left to live conditional on their age (someone at 84 has, on average, 2 years left, say). These are the estimates given in actuarial ‘death tables’. So how do we make sense of the notion?


- If we are not careful, the deprivation account of the badness of death runs into the same Epicurean problem that we were trying to avoid. If we say that death is bad because of the goods that it deprives us of, we may be saying that you would be better off with the things you would have experienced than you would be if you were dead, but of course it’s not like anything to be dead, so this comparison is nonsensical.
- When we have solved the problem of why death is bad for the person for whom it is bad, we are faced with the question of when it is bad. There is discussion of this in the notes. Feldman also brings up the problem of why, if the deprivation account is correct, we do not lament that we weren’t born earlier.
- Some commenters think that Epicurus’ view about death is inextricably tied to his hedonism (in that having no positive or negative mental states means that death is neutral, and thus not bad for you). But this only holds true if your future life is neutral in expectation.
- Nagel (see above) has a view in which death is bad in much the same way having “nasty rumours spread about you is bad”.
- We might distinguish between death and you dying. It may be that it makes no sense to say that ‘smoking’ is bad for Greg, but it is bad for Greg that he smokes, which is a potentially relevant distinction in the death case.
- We can also draw the distinction between things being intrinsically bad for someone (pure suffering) and all things considered bad – i.e. worse than the counterfactual, while possibly being intrinsically neutral, like moving country.
- Feldman talks about the all things considered variety of something being worse for you as being predicated on how things would have gone in nearby possible worlds – e.g. how do I fare if my plane is hijacked, compared to the nearby possible worlds in which it isn’t. He assumes that there exists a unique nearest possible world. I’m not sure how much the argument rests upon this assumption, but I think it’s unjustified.
• Suppose that Ivan is shot and killed by Dmitry, but, if he had not been shot, he would have died seconds later by being shot by Boris. Can we say that being shot by Dmitry was bad for Ivan? We can resolve this paradox by decomposing the badness of Ivan’s being shot, something like ‘Ivan died young’, ‘Ivan was shot’, ‘Ivan was shot by Dmitry’. In this case, the last component does not matter, while in other cases that component would matter. Parfit talks about this, giving an example of one of his “mistakes in moral mathematics”. If we suppose that five men in a firing squad all fire upon a prisoner, then the ‘marginal product’ of any one man is zero, and yet it would be strange to absolve any one of moral responsibility for the prisoner’s death.

Montaigne, Michel de. “That to Philosophize is to Learn to Die” In The Complete Essays
• Cicero says “that to study philosophy is nothing but to prepare one’s self to die”. I like this quote.
• Montaigne seems to have a similar reason for thinking about death that I do, namely that, by considering death and philosophising about it, it won’t creep up on you. The fear of death is thus smeared out over a lifetime, so that it is manageable.
• We are fearful of an infinite variety of possible deaths, yet only experience one.

Swift, Jonathan. Gulliver’s Travels, Part III, chapter 10
• In the book, the struldbrugs are a race of immortals. Gulliver thinks at first that these people are blessed, but has a dialogue in which he learns about the curse of immortality.

• Some philosophers, like the existentialists, have argued that the way in which death gives meaning to life is because of the fear of death, but Williams here is arguing that it is more the significance of the fact that we die that gives us meaning.
• It’s pretty interesting that there are such temperamental differences in the attitudes (at least, the stated ones) toward death. There are those who are comforted by believing that there is life after death, and those who are comforted by believing that death is the absolute end.
• Williams brings up the Nagel objection to the Epicurean argument, which is that one does not need to know about something for it to be bad for them. Death is bad for you in much the same way having someone gossip negatively behind your back is.
  o You could actually move Nagel’s view in a more consequentialist direction by taking the expected value of actions as the benefit of harm of those actions. Thus you are in fact harmed by someone gossiping about you if there is a 10% chance that you will find out and be emotionally hurt, but if that probability is 0%, then you are not harmed (nor benefited in the opposite case). I can imagine this running into problems with determinism however.
• Williams argues that it is pretty hard to strike the balance if you are immortal between enough memory change so as to preserve the identity by which you would want to survive, but not so much stasis that you fall into the pitfalls of boredom and insanity.
• Williams thinks that the total loss of character involved in immortality would correspond to a total loss in individuality, and so immortality loses its initial appeal.
Review of the Course

I enjoyed this course. My main complaint was probably Kagan was repetitive at times. It certainly seems like there are many other topics that could have been included in a course of this size (although it may be better to just make the same points and shorten the course). These include population ethics, the ethics of heredity (do I deserve my father’s wealth when he dies?), preference satisfaction (am I obligated to carry out the wishes of a dead person?) and the strange views of the transhumanist/cryonics community. These are people who want to freeze their bodies after their die so that some future medical technology can thaw them out. I would like to see a Kagan-style investigation of the philosophical significance of this.

I think too much time was spent on souls and discussing arguments and counter-arguments about whether or not there is an afterlife. I did enjoy this discussion and it is certainly relevant, but there was an elephant in the room: that these hypotheses were getting more attention than they deserved because many of the students are religious, and America is very religious in general. If we replaced the soul hypothesis with something with a similar level of philosophical grounding, there is no way it would get such coverage in a philosophy class.

My favourite part was the discussion of the Phaedo, which surprised me. I didn’t have much exposure to ancient philosophy before this.

On the major points I am with Kagan, in that I think that it is good that we die, and that physicalism is true. However, it surprised me that Kagan did not take more time to discuss the existence of phenomenological consciousness as a refutation to physicalism. There is nothing whatsoever mysterious about why people just die, but there is something mysterious about how a purely physical system would be conscious. And insofar as consciousness is discussed, it would have been helpful to have made some distinctions. The soul is sometimes seen as synonymous with the ‘self’, sometimes seen as synonymous with ‘that which persists after death’, and sometimes synonymous with phenomenological consciousness. The course felt to me vague on this point.