PART I: GOD

ESSAY 1: The Ontological Argument, Saint Anselm

Anselm of Canterbury was the first to attempt an ontological argument for God's existence.

Theologian and philosopher Anselm of Canterbury (1033–1109) proposed an ontological argument in the second and third chapters of his *Proslogion*. Anselm's argument was not presented in order to prove God's existence; rather, *Proslogion* was a work of meditation in which he documented how the idea of God became self-evident to him.

In Chapter 2 of the *Proslogion*, Anselm defined God as a "being than which no greater can be conceived". He suggested that even "the fool" can understand this concept, and this understanding itself causes the being to exist in the mind. The concept must exist either only in our mind, or in both our mind and in reality. If such a being exists only in our mind, then a greater being—that which exists in the mind and in reality—can be conceived. Therefore, if we can conceive of a being of which nothing greater can be conceived, it must exist in reality. Thus, a being than which nothing greater could be conceived, which Anselm defined as God, must exist in reality.

Anselm's argument in Chapter 2 can be summarized as follows:

1. Our understanding of God is a being than which no greater can be conceived.
2. The idea of God exists in the mind.
3. A being which exists both in the mind and in reality is greater than a being that exists only in the mind.
4. If God only exists in the mind, then we can conceive of a greater being—that which exists in reality.
5. We cannot be imagining something that is greater than God.
6. Therefore, God exists.

In Chapter 3, Anselm presented the notion of a being that cannot be conceived to not exist. He argued that if something can be conceived to not exist, then something greater can be conceived. Consequently, a thing than which nothing greater can be conceived cannot be conceived to not exist and so it must exist. This can be read as a restatement of the argument in Chapter 2, although Norman Malcolm believed it to be a different, stronger argument.

ESSAY 2: In Behalf of the Fool, Gaunilo

Gaunilo criticised Anselm's argument by employing the same reasoning, via *reductio ad absurdum*, to "prove" the existence of the mythical "Lost Island", the greatest or most perfect island conceivable: if the island of which we are thinking does not exist, it cannot be the greatest conceivable island, for, to be the greatest conceivable island, it would have to exist, as any existent island would be greater than an imaginary one. This, of course, is merely a direct application of Anselm's own premise that existence is a perfection. Since we can conceive of this greatest or most perfect conceivable island, it must, by Anselm's way of thinking, exist. While this argument is absurd, Gaunilo claims that it is no more so than Anselm's.

Philosophers often attempt to prove the ontological argument wrong by comparing Anselm's with Gaunilo's. The former runs thus:

1. God is that being than which no greater can be conceived.
2. It is greater to exist in reality than merely as an idea.
3. If God does not exist, we can conceive of an even greater being, *that is* one that *does* exist.
4. Therefore, God must indeed exist in reality.
5. Therefore, He exists.

Gaunilo's argument runs along the same lines:

1. The Lost Island is that than which no greater can be conceived.
2. It is greater to exist in reality than merely as an idea.
3. If the Lost Island does not exist, one can conceive of an even greater island, that is one that does exist.
4. Therefore, the Lost Island exists in reality.

If one of these arguments is sound, it has been asserted, they must both be sound. By Gaunilo's reckoning, however, one (and, therefore, the other, too) is unsound. The Lost Island does not exist, so there is something wrong with the logic that proves that it does. Because the argument proves true in one case that which is patently false (the Lost Island), it is fair to ask whether it may fairly be regarded as proving true

the other case. The fact that there is no perfect island is put forth by Gaunilo as showing that Anselm's argument for God's existence is flawed.

Such objections are called overload objections: they do not claim to show where or how the argument goes wrong; they merely argue that, if it is unsound in one application, it is unsound in all others. Simply put, they are arguments that would overload the world with an indefinitely large number of things, like perfect islands.

ESSAY 3: Why the Ontological Argument Fails, William L. Rowe

Rowe begins by making a distinction between two types of arguments for the existence of God that might be given. The first such sort is an *a posteriori* argument. The defining feature of arguments of this kind is that they rely on actual experiences that we have had of the world. Another kind of argument, an *a priori* argument, relies on no such experiences. These arguments could be reasoned through even if we had had no interaction with the world; Rowe points out that the Ontological Argument is an example of this second type of argument. The argument is supposed to rely only on an analysis of the concept of God in order to demonstrate that such a being must exist.

In §I, Rowe attempts to clarify important terms and ideas in Anselms argument, and ultimately states what he takes the key idea to be in the Ontological Argument. This idea is that, for Anselm, existing in reality is great-making quality. Rowe reminds us that Anselm distinguishes between *existence in the understanding*, and *existence in reality*. The idea is that for a thing to exist in reality is for it to exist in the world in the way that we ordinarily understand objects like desks, chairs, cats, planets, and electrons to exist. Existing in the understanding, though, is much different. Although unicorns are not real objects in the world, they exist in the understanding in the sense that we have an idea of what a unicorn; it exists, we might say, in our minds. What Rowe is identifying as Anselms key idea is that, for any being that exists only in the understanding, but *could have* existed in reality, that being could have been greater than it is. Following this, Rowe formulates in §II what he takes to be Anselms argument.

In the final sections of the selection, Rowe discusses and evaluates several kinds of objections that have been raised against the Ontological Argument. He first discusses Gaunilos objection involving the example of the island, but ultimately defends Anselm against this objection, identifying two problems in Gaunilos reasoning. The second objection contends that Anselms argument is flawed because it treats existence as a predicate, but it in fact is not. Rowe points out that a crucial point to this objection is the claim that when we say that a being has a certain property, we are thereby implicitly asserting that that being exists. If this were true, then it would not make sense to treat existence as a predicate, as we could only ascribe it to something if we assumed that the thing existed. Rowe contends that we do not treat predication in this way

at all, and that the objection therefore loses its force. The last objection that Rowe considers in §III contends that the type of God that Anselm has in mind is actually not possible. Again, Rowe demonstrates that the reasoning behind this objection is flawed.

In §IV, Rowe addresses a final kind of criticism of the Ontological Argument. In the first place, he examines Anselms concept of God, and examines whether anything concerning Gods existence follows from that concept all on its own. He argues that although it does follow immediately from that definition that no non-existing thing that could have existed could be God, it does not follow that God actually exists. After all, the concept cannot itself demonstrate that something *instantiates* that concept; it remains to be seen whether some actually existing thing does in fact have the properties that Anselm has in mind. The analysis cannot stop here, though, as Rowe notes. The criticism loses force once we pay heed to Anselms claim that God is a possible thing. Any possible thing will be either existing or non-existing; if it really is true that Anselms concept of God rules out the possibility of God being a non-existing thing, then it follow from Gods possibility that God must be an existing thing. In light of this, it looks as if it really must be the case that God exists in reality. Rowe points out, though, that we need not be so quick to grant to Anselm that God is a possible thing. This is because, given his understanding of God and his assertion that existence is a great-making feature, allowing that God is possible is just equivalent to allowing that God is an actually existing thing. And surely this is too much to ask. Rowe allows that Anselms argument succeeds only if there is already some existing thing that is maximally perfect, but denies that the argument successfully shows that such being must exist.

ESSAY 4: The Argument from Design, William Paley

Paley's argument is built mainly around anatomy and natural history. "For my part", he says, "I take my stand in human anatomy"; elsewhere he insists upon "the necessity, in each particular case, of an intelligent designing mind for the contriving and determining of the forms which organized bodies bear". In making his argument, Paley employed a wide variety of metaphors and analogies. Perhaps the most famous is his analogy between a watch and the world. Historians, philosophers and theologians often call this the Watchmaker analogy.

ESSAY 5: The Problem of Evil, John Hick

A Soul-Making Theodicy

One very important type of theodicy, championed especially by John Hick, involves the idea that the evils that the world contains can be seen to be justified if one views the world as designed by God as an environment in which people, through their free choices can undergo spiritual growth that will ultimately fit them for communion with God:

The value-judgement that is implicitly being invoked here is that one who has attained to goodness by meeting and eventually mastering temptation, and thus by rightly making responsibly choices in concrete situations, is good in a richer and more valuable sense than would be one created ab initio in a state either of innocence or of virtue. In the former case, which is that of the actual moral achievements of mankind, the individual's goodness has within it the strength of temptations overcome, a stability based upon an accumulation of right choices, and a positive and responsible character that comes from the investment of costly personal effort. (1977, 255-6)

Hick's basic suggestion, then, is that soul-making is a great good, that God would therefore be justified in designing a world with that purpose in mind, that our world is very well designed in that regard, and thus that, if one views evil as a problem, it is because one mistakenly thinks that the world ought, instead, to be a hedonistic paradise.

Is this theodicy satisfactory? There are a number of reasons for holding that it is not. First, what about the horrendous suffering that people undergo, either at the hands of others—as in the Holocaust—or because of terminal illnesses such as cancer? One writer—Eleonore Stump—has suggested that the terrible suffering that many people undergo at the end of their lives, in cases where it cannot be alleviated, is to be viewed as suffering that has been ordained by God for the spiritual health of the individual in question. (1993b, 349). But, given that it does not seem to be true that terrible terminal illnesses more commonly fall upon those in bad spiritual health than upon those of good character, let alone that they fall only upon the former, this 'spiritual chemotherapy' view seems quite hopeless. More generally, there seems to be no reason at all why a world must contain horrendous suffering if it is to provide a good environment for the development of character in response to challenges and temptations.

Secondly, and is illustrated by the weakness of Hick's own discussion (1977, 309-17), a soul-making theodicy provides no justification for the existence of any animal pain, let alone for a world where predation is not only present but a major feature of non-human animal life. The world could perfectly well have contained only human persons, or only human person plus herbivores.

Thirdly, the soul-making theodicy provides no account either of the suffering that young, innocent children endure, either because of terrible diseases, or at the hands of adults. For here, as in the case of animals, there is no soul-making purpose that is served.

Finally, if one's purpose were to create a world that would be a good place for soul-making, would our earth count as a job well done? It is very hard to see that it would. Some people die young, before they have had any chance at all to master temptations, to respond to challenges, and to develop morally. Others endure suffering so great that it is virtually impossible for them to develop those moral traits that involve relationships with others. Still others enjoy lives of ease and luxury where there is virtually nothing that challenges them to undergo moral growth.

PART II: FREE WILL

ESSAY 6: Freedom and Necessity, A. J. Ayer

A. The Significance of the Problem of Free Will

1. People have free will when they are in control of their actions. People who are not in control of their actions are not morally responsible for what they do. So treating people as moral agents worthy of praise or blame for what they do requires that they have free will.
2. Cognitive science proposes to explain all human action as the result of the activity of the brain. If you adopt this naturalist point of view, then it seems that there is a direct conflict with what is required for free will. If all I ever do is the result of brain events subject to the laws of nature, then those laws are responsible for my actions, not my will; and if my will is not responsible, I am not responsible for my actions.
3. There seems to be a fundamental conflict here between morality and naturalism. Morality requires freedom of the will while naturalism appears to require determination by natural law. It seems that something has to give - morality or cognitive science.

B. The Shape of the Problem of Free Will

1. **Libertarians** believe: Human actions are the result of free will.
2. **Determinists** believe: Human actions are determined by natural law.
3. It would seem that Libertarian and Determinist ideas are incompatible. However, **Compatibilists** believe: Free will is compatible with determinism. They hope to find some way to preserve both morality and naturalism at the same time.

C. Strategies for Solving the Problem of Free Will

1. Deny Libertarianism. One tactic is simply to give up on free will. Skinner takes this view in *Beyond Freedom and Dignity*. Although it seems to us as though we have free will, we are mistaken. Our belief in free will is due to our ignorance about how we work. Since we do not understand the natural laws that fix our actions, we assume that we (and not they) are really in control. But we are fooling ourselves. Free will is an illusion.

This idea is supported by two kinds of evidence from cognitive psychology. First, the left hemispheres of split brain subjects **confabulate** (make up stories after the fact) about why they were in control of actions of their right hemispheres. Maybe normal brains confabulate freedom in the same way. Second, timing studies of decision making suggest that we are conscious of our making a decision only *after* the action decided on is well underway. Our conscious mind may confabulate stories about why the body is doing what it is doing, without really being in control at all.

2. Deny Determinism. Science may not have an explanation for everything, especially when it comes to what people do. Some events may not be controlled by physical laws. Ayer points out, however, that denying determinism does not get us out of the problem of free will. Suppose that some events in the physical world are not determined. Then if you are a naturalist, they must be due to chance. Now chance actions are still not in *our* control. We are no more responsible for accidental events than we are responsible for ones that are determined. So the problem is to find room for freedom in a natural world where events are determined by chance and natural law. Chance is not what we want for freedom. What we want is control *by us* . So what it seems that one needs to deny is that free actions are controled by either natural law or chance. So all our free actions cannot be explained by any conceivable science. To put it another way: free actions are miracles.
3. Compatibilism. Denying Libertarianism and denying natural explanation seem harsh conclusions. Many philosophers have hoped to resolve the free will problem by showing how the two views are both right. Compatibilists believe that freedom is possible even if all events are the result of natural law and/or chance. Ayer discusses and objects to one brand of compatibilism that asserts that freedom is the consciousness of necessity. This view says that we are free when we come to accept our destiny. The problem with this, says Ayer, is that freedom is defined in a strange way. What is needed is a way to explain why freedom *as ordinarily understood* is compatible with natural law and chance.

D. Ayer's Compatibilist Solution to the Problem of Free Will

1. Ayer begins his solution by asking: what is the opposite of freedom? People think the opposite of being free is being determined or caused. Ayer suggests that is wrong. The opposite of freedom is *constraint* not cause. Think about situations where we would say that a person's actions weren't free - where the person was not in control of his/her actions. In cases like this, the person is threatened, or hypnotized, or brainwashed. In general, a person is under constraint when their actions are not determined by their rational decision making abilities.
2. A person is free of constraint when he/she would have acted otherwise if he/she had decided otherwise. Such a person is not acting from a compulsion or an addiction or reacting to threat from another person. Being free in this way is perfectly compatible with the ex-

istence of causal mechanisms in the brain that would allow us to predict and explain every one of our actions. In fact causal connections between my decisions and my actions are a requirement for free action.

3. It is a mistake to think that freedom requires that my actions not be caused, or that they not be predicted or explained. Part of the problem may be that by taking the metaphor of causality too seriously we confuse causality with constraint. We think a cause is something like a person threatening us or making us do something, forgetting that causal mechanisms are actually the foundation of our abilities to act freely.
4. It might be objected that we can never be free if all our actions are determined at the beginning of the universe. How can I be free if molecular action will inevitably fix what I do? Aren't I then a "helpless prisoner of fate" (Ayer, p. 22). Ayer replies that determinism does not entail that I am a helpless prisoner. What matters to my freedom is whether I exercise my decision making abilities to the fullest - whether if I had decided otherwise I would have acted otherwise. If we discover that these conditions are met, then my action is no less free for being ruled by natural law.

ESSAY 7: Human Freedom and the Self, Roderick M. Chisholm

Chisholm argues that determinism is incompatible with free will, but that nonetheless humans have free will. He rejects compatibilist arguments offered by many philosophers (e.g., Ayer and Stace from this volume). He argues that a deterministic universe, where all events, including human actions, proceed from prior events without the possibility that they would proceed differently than they do precludes the possibility that humans are responsible for their actions.

Chisholm rejects the following sort of argument for thinking that free will and determinism are compatible. Having free will simply means that (a) one could have done otherwise, which in turn means precisely the same thing as (b) if one had chosen to do otherwise, then one would have done otherwise. Statement (b) is compatible with determinism, and because (a) and (b) mean the same thing, (a) is compatible with determinism as well. Thus, free will is compatible with determinism. Chisholm, however, argues that (a) and (b) are *not* equivalent, and that (a) can follow from (b) only if it is also true that (c) one could have chosen to do otherwise. Whether or not one can really choose otherwise, then, is the crux of the free will debate.

Having articulated the question in this way, Chisholm points out just how extraordinary the notion of free will is. For an agent's actions to be chosen freely, they must not be caused by events, they must not occur by mere chance, and they must not be uncaused (the three of which seem to exhaust the possible explanations for most phenomena). Rather, the agent must cause them. Moreover, the agent must not have been in turn caused to act for her action to be free. That is, an action is the result of free will only if an agent acts because of a choice that is not itself caused by other events.

Chisholm admits that it is hard to see how an event can be caused by an agent and at the same time not be caused by other events. That is, one might view an event as resulting from a series of other events, some of which involve an agent's action, and all of which could not have occurred otherwise. On this picture, saying that the agent caused the event does not add anything to a description of the causal events and resulting events; saying that an agent causes an action is just another way to describe event-causation. However, Chisholm thinks that this is a mistake. He suggests that the reason it seems plausible to reduce agent-causation to event-causation is our poor understanding of causation it-

self. In fact, it is difficult to see what the concept of *causation* adds to a description of events that follow one another in precisely the same way it is hard to see what the idea of agent-causation adds to event-causation. Thus, Chisholm finds the possibility of agent-causation plausible.

PART III: PERSONAL IDENTITY

ESSAY 8: “Of Identity and Diversity” from an Essay Concerning Human Understanding", John Locke

人類理解論

1. Wherein identity consists.
   That, therefore, that had one beginning, is the same thing; and that which had a different beginning in time and place from that, is not the same, but diverse.
2. Identity of substances.
   three sorts of substances: God, finite intelligences, bodies.
   identity of modes and relations.
   All other things being but modes or relations ultimately terminated in substances, …only as to things whose existence is in succession, such as are the actions of finite beings, …
3. Principium Individuationis (Principle of individuation)
   For in them the variation of great parcels of matter alters not the identity: an oak growing from a plant to a great tree, and then lopped, is still the same oak; and a colt grown up to a horse, sometimes fat, sometimes lean, is all the while the same horse: though, in both cases, there may be a manifest change of the parts; so that truly they are not either of them the same masses of matter, though they be truly one of them the same oak, and the other the same horse. The reason whereof is, that, in these two cases–a mass of matter and a living body–identity is not applied to the same thing.
4. Identity of vegetables.
5. Identity of animals.
6. Identity of man.
   This also shows wherein the identity of the same man consists; viz. in nothing but a participation of the same continued life, by constantly fleeting particles of matter, in succession vitally united to the same organized body. He that shall place the identity of man in anything else, but, like that of other animals,

in one fitly organized body, taken in any one instant, and from thence continued, under one organization of life, in several successively fleeting particles of matter united to it, will find it hard to make an embryo, one of years, mad and sober, the same man, by any supposition, that will not make it possible for Seth, Ismael, Socrates, Pilate, St. Austin, and Caesar Borgia, to be the same man. For if the identity of soul alone makes the same man; and there be nothing in the nature of matter why the same individual spirit may not be united to different bodies, it will be possible that those men, living in distant ages, and of different tempered, may have been the same man: which way of speaking must be from a very strange use of the word man, applied to an idea out of which body and shape are excluded. And that way of speaking would agree yet worse with the notions of those philosophers who allow of transmigration, and are of opinion that the souls of men may, for their miscarriages, be detoured into the bodies of beasts, as fit habitations, with organs suited to the satisfaction of their brutal inclinations. But yet I think nobody, could he be sure that the soul of Heliogabalus were in one of his hogs, would yet say that hog were a man or Heliogabalus.

"Soul alone cannot make a man. Body counts."

7. Identity suited to the idea.

   It is not therefore unity of substance that comprehends all sorts of identity, or will determine it in every case; but to conceive and judge of it aright, we must consider what idea the word it is applied to stands for: it being one thing to be the same substance, another the same man, and a third the same person, if person, man, and substance, are three names standing for three different ideas; for such as is the idea belonging to that name, such must be the identity; which, if had been a little more carefully attended to, would possible have prevented a great deal of that confusion which often occurs about this matter, with no small seeming difficulties, especially concerning personal identity, which therefore we shall in the next place a little consider.

8. Same man.

   An animal is a living organized body; and consequently the same animal, as we have observed, is the same continued life communicated to different particles of matter, as they happen successively to be united to that organized living body.

   …

   Since I think I may be confident, that, whoever should see a creature of his own shape or make, though it had no more reason all its life than a cat or a parrot, would call him still a man; or whoever should hear a cat or a parrot discourse, reason, and philosophize, would call or think it nothing but a cat or a parrot; and say, the one was a dull irrational man, and the other is a very intelligent rational parrot. … For I presume it is not the idea of a thinking or rational being alone that makes the idea of a man, the same successive body not shifted all at once, must, as well as the same immaterial spirit, go to the making of the same man.

9. Personal identity.

   This being premised, to find wherein personal identity consist, we must consider what person stands for; which, I think, is a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness

which is inseparable from thinking, and, as it seems to me, essential to it: it being impossible for any one to perceive without perceiving that he does perceive. When we see, hear, smell, taste, feel, meditate, or will anything, we know that we do so. Thus it is always as to our present sensations and perceptions: and by this every one is to himself that which he calls self: it not being considered, in this case, whether the same self be continued in the same or divers substances. For, since consciousness always accompanies thinking, and it is that which makes every one to be what he calls self, and thereby distinguishes himself from all other thinking things, in this alone consists personal identity, i.e. the sameness of a rational being: and as far as this consciousness can be extended backwards to any past action or though, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it, that that action was done.

10. Consciousness makes personal identity.
    … I say, in all these cases, our consciousness being interrupted, and we losing the sight of our past selves, doubts are raised whether we are the same thinking thing, i.e. the same substance or no. Which, however reasonable or unreasonable, concerns not personal identity at all. … For, it being the same consciousness that makes a man be himself to himself, personal identity depends on that only, whether it be annexed solely to one individual substance, or can be continued in a succession of several substances. For as far as any intelligent being can repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present action; so far it is the same personal self. For it is by the consciousness it has of its present thoughts and actions, that it is self to itself now, and so will be the same self, as far as the same consciousness can extend to actions past or to come and would be by distance of time, or change of substance, no more two persons, than a man be two men by wearing other clothes to-day than he did yesterday, with a long or a short sleep between: the same consciousness uniting those distant actions into the same person, whatever substances contributed to their production.
11. Personal identity in change of substances.
    … we have some kind of evidence in our very bodies, all whose particles, whilst vitally united to this same thinking conscious self, so that we feel when they are touched, and are affected by, and conscious of good or harm that happens to them, as a part of ourselves; i.e. of our thinking conscious self. Thus, the limbs of his body are to every one a part of Himself; he sympathizes and is concerned for them. Cut off a hand, and thereby separate it from that consciousness he had of its heat, cold, and other affections, and it is then no longer a part of that which is himself, any more than the remotest part of matter. Thus, we see the substance whereof personal identity; there being no question about the same person, though the limbs which but now were a part of it, be cut off.
12. Whether in the change of thinking substances. "thinking changed, still being the same person?"
    First, this can be no question at all to those who place thought in a purely material animal constitution, void of an immaterial substance. … And therefore those who place thinking in an immaterial substance only, before they can come to deal with these men, must show why personal identity cannot be preserved in the change of immaterial substances, or variety of particular immaterial substances, as well as animal identity is preserved in the change of material substances, or variety of particular bodies: unless they will say, it is one immaterial spirit that makes the same life in brutes, as it is one immaterial

spirit that makes the same person in men; which the Cartesians at least will not admit, for fear of making brutes thinking things too.

13. continued.

first part of the question. … I answer, that cannot be resolved but by the those who know what kind of substances they are that do think; and whether the consciousness of past actions can be transferred from one thinking substance to another. … it must be allowed, that, if the same consciousness (which, as has been shown, is quite a different thing from the same numerical figure or motion in body) can be transferred from one thinking substance to another, it will be possible that two things substances may make but one person. For the same consciousness being preserved, whether in the same or different substances, the personal identity is preserved.

14. continued.

second part of the question. Whether the same immaterial substance remaining, there may be two distinct persons; which question seems to me to be built on this,—Whether the same immaterial being, being conscious of the action of its past duration, may be wholly stripped of all the consciousness of its past existence, and lose it beyond the power of ever retrieving it again: and so as it were beginning a new account from a new period, have a consciousness that cannot reach beyond this new state. … So that this consciousness, not reaching to any of the actions of either of those men, he is no more one self with either of them than if the soul or immaterial spirit that now informs him had been created, and began to exist, when it began to inform his present body; though it were never so true, that the same spirit that informed Nestor's or Thersites' body were numerically the same that now informs his. For this would no more make him the same person with Nestor, than if some of the particles of matter that were once a part of Nestor were now a part of this man; the same immaterial substance, without the same consciousness, no more making the same person, by being united to any body, than the same particle of matter, without consciousness, united to any body, makes the same person. But let him once find himself conscious of any of the actions of Nestor, he then finds himself the same person with Nestor.

15. continued.

And thus may we be able, without any difficulty, to conceive the same person at the resurrection, though in a body not exactly in make or parts the same which he had here,—the same consciousness going along with the soul that inhabits it. But yet the soul alone, in the change of bodies, would scare to any one but to him that makes the soul the man, be enough to make the same man. … The body too goes to the making the man, and would, i guess, to everybody determine the man in this case, wherein the soul, with all its princely thoughts about it, would not make another man: but he would be the same cobbler to every one besides himself. … But yet, when we will inquire what makes the same spirit, man, or person, we must fix the ideas of spirit, man, or person in our minds; and having resolved with ourselves what we mean by them, it will not be hard to determine, in either of them, or the like, when it is the same, and when not.

16. Consciousness makes the same person.

But though the same immaterial substance or soul does not alone, wherever it be, and in whatsoever state, make the same man; yet it is plain, conscious-

ness, as far as ever it can be extended–should it be to ages past–unites existences and actions very remote in time into the same person, as well as it does the existences and actions of the immediately preceding moment: so that whatever has the consciousness of present and past actions, is the same person to whom they both belong. …

17. Self depends on consciousness.
    Self is that conscious thinking thing,–whatever substance made up of, (whether spiritual or material, simple or compounded, it matters not)–which is sensible or conscious of pleasure and pain, capable of happiness or misery, and so is concerned for itself, as far as that consciousness extends. …
18. Object of reward and punishment.
19. continued.
    This may show us wherein personal identity consists: not in the identity of substance, but, as I have said, in the identity of substance, but, as I have said, in the identity of consciousness, wherein if Socrates and the present mayor of Quinborough agree, they are the same person: if the same Socrates waking and sleeping do not partake of the same consciousness, Socrates waking and sleeping is not the same person. …
20. continued.
    … Suppose I wholly lose the memory of some parts of my life, beyond a possibility of retrieving them, so that perhaps I shall never be conscious of them again; yet am I not the same person that did those actions, had those thoughts that I once was conscious of, though I have now forgot them? To which I answer, that we must here take notice what the word I is applied to; which, in this case, is the man only. And the same man being presumed to stand for the same person, I is easily here supposed to stand also for the same person. But if it be possible for the same man to have distinct incommunicable consciousness at different times, it is past doubt the same man would at different times make different persons; which, we see, is the sense of mankind in the solemnest declaration of their opinions, human laws not punishing the mad man for the sober man's action, nor the sober man for what the mad man did,–thereby making them two persons: which is somewhat explained by our ways of speaking in English when we say such an one is "not himself," or is "beside himself"; in which phrases it is insinuated, as if those who now, or at least first used them, thought that self was changed; the selfsame person was no longer in that man.
21. Difference between identity of man and of person.
    … we must consider what is meant by Socrates, or the same individual man. First, it must be either the same individual, immaterial, thinking substance; in short, the same numerical soul, and nothing else. Secondly, or the same animal, without any regard to an immaterial soul. Thirdly, or the same immaterial spirit united to the same animal. … personal identity can by us be placed in nothing but consciousness. …
22. continued.
23. Consciousness alone makes self.
    Nothing but consciousness can unite remote existences into the same person: the identity of substance will not do it; for whatever substance there is, however

framed, without consciousness there is no person: and a carcass may be a person, as well as any sort of substance be so, without consciousness. … So that self is not determined by identity or diversity of substance, which it cannot be sure of, but only by identity of consciousness.

24. continued.
25. continued.
26. "Person" a forensic term.

    … This personality extends itself beyond present existence to what is past, only by consciousness,—whereby it becomes concerned and accountable; owns and imputes to itself past actions, just upon the same ground and for the same reason as it does the present. …
27. The difficulty from ill use of names.

    To conclude: Whatever substance begins to exist, it must, during its existence, necessarily be the same: whatever compositions of substances begin to exist, during the union of those substances, the concrete must be the same: whatsoever mode begins to exist, during its existence it is the same: and so if the composition be of distinct substances and different modes, the same rule holds. Whereby it will appear, that the difficulty or obscurity that has been about this matter rather rises from the names ill-used, than from any obscurity in things themselves. For whatever makes the specific idea to which the name is applied, if that idea be steadily kept to, the distinction of anything into the same and divers will easily be conceived, and there e can arise no doubt about it.
28. Continued existence makes identity.

ESSAY 9: Of Mr. Locke's Account of Our Personal Identity, Thomas Reid

…

Identity, as was observed, supposes the continued existence of the being of which it is affirmed, and therefore can be applied only to things which have a continued existence. While any being continues to exist, it is the same being; but two beings which have a different beginning or a different ending of their existence cannot possibly be the same. To this, I think, Mr. Locke agrees.

He observes, very justly, that, to know what is meant by the same person, we must consider what the word person stands for; and he defines a person to be an intelligent being, endowed with reason and with consciousness, which last he thinks inseparable from thought.

From this definition of a person, it must necessarily follow, that, while the intelligent being continues to exist and to be intelligent, it must be the same person. To say that the intelligent being is the person, and yet that the person ceases to exist while the intelligent being continues, or that the person continues while the intelligent being ceases to exist, is to my apprehension a manifest contradiction.

…

Mr. Locke tells us, however, "that personal identity, that is, the sameness of a rational being, consists in consciousness alone, and, as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person. So that whatever has the consciousness of present and past actions is the same person to whom they belong."

… some strange consequences, … Such as, that if the same consciousness can be transferred from one intelligent being to another, which he thinks we cannot show to be impossible, then two or twenty intelligent beings may be the same person. And if the intelligent being may lose the consciousness of the actions done by him, which surely is possible, then he is not person that did those actions; so that one intelligent being may be two or twenty different persons, if he shall so often lose the consciousness of his former actions.

… another consequence of this doctrine, which follows no less necessarily, though Mr. Locke probably did not see it. It is, *that a man be, and at the same time not be, the person that did a particular action*.

… A = B, B = C, but A != C

Firstly, … It is impossible to understand the meaning of this, unless by consciousness be meant memory, the only faculty by which we have an immediate knowledge of our past actions.

… If a man can be conscious of what he did twenty years or twenty minutes ago, there is no use for memory, nor ought we allow that there is any such faculty. The faculties of consciousness and memory are chiefly distinguished by this, that the first is an immediate knowledge of the present, the second an immediate knowledge of the past.

Secondly, … personal identity is confounded with the evidence which we have of our personal identity.

…The only evidence I have that I am the identical person who did such actions is, that I remember distinctly I did them; or, as Mr. Locke expresses it, I am conscious I did them. To infer from this, that personal identity consists in consciousness, is an argument which, if it had any force, would prove the identity of a stolen horse to consist solely in similitude.

Thirdly, is it not strange that the sameness or identity of a person should consists in a thing which is continually changing, and is not any two minutes the same?

… Identity can only be affirmed of things which have a continued existence. Consciousness, and every kind of thought, are transient and momentary, and have no continued existence; and, therefore, if personal identity consisted in consciousness, it would certainly follow, that no man is the same person any two moments of his life; and as the right and justice of reward and punishment are founded on personal identity, no man could be responsible for his actions.

… Fourthly, there are many expressions used by Mr. Locke, in speaking of personal identity, which to me are altogether unintelligible, unless we suppose that he confounded that sameness or identity which we ascribe to an individual with the identity which, in common discourse, is often ascribed to many individuals of the same person.

… this sameness can only mean similarity, or sameness of kind…

If our personal identity consists in consciousness, as this consciousness cannot be the same individually any two moments, but only of the same kind, it would follow, that we are not for any two moments the same individual persons, but the same kind of persons.

As our consciousness sometimes ceases to exist, as in sound sleep, our personal identity must cease with it. Mr. Locke allows, that the same thing cannot have two beginnings of existence, so that our identity would be irrecoverably gone every time we ceased to think, if it was but for a moment.

Basically an objection to Locke's Memory Criterion.

ESSAY 10: Personal Identity from Reasons and Persons, Derek Parfit

What we believe ourselves to be

Qualitative and numerical identity

Numerical identity: X and Y are numerically identical if and only if they are one and the same thing. (e.g. Bruce Wayne and Batman)

Qualitative identity: X and Y are qualitatively identical if and only if they have exactly the same properties. (e.g. identical watches, Big Macs)

The physical criterion of personal identity

(1) What is the nature of a person?

Answer: to be a person, a being must be self-conscious, aware of its identity and its continued existence over time.

(2) What makes a person at two different times one and the same person? What is necessarily involved in the continued existence of each person over time?

Answer: X today is one and the same person as Y at some past time if and only if… Such an answer states the necessary and sufficient conditions for personal identity over time.

or (2) is "similar" to ask (3) What is in fact involved in the continued existence of each person over time?

Answer: the answer to (2) is only part of the answer to (3) since our continued existence has features that are not necessary.

For most physical objects, in the *standard view*, the criterion of identity over time is the *spatio-temporal physical continuity* of this object.

Some kinds of thing continue to exist even though their physical continuity involves great changes.

Another complication again concerns the relation between a complex thing and the various parts of which it is composed. It is true of some of these things, though not true of all, that their continued existence need not involve the continued existence of their components.

On this view, what makes me the same person over time is that I have the same brain and body. The criterion of my identity over time–or what this identity involves–is the physical continuity, over time, of my brain and a body. I shall continue to exist if and only if this particular brain and body continue both to exist and to be the brain an body of a living person.

Simplified version:

The Physical Criterion (1) What is necessary is not the continued existence of the whole body, but the continued existence of enough of the brain to be the brain of a living person. X today is one and the same person as Y at some past time if and only if (2) enough of Y's brain continued to exist, and is now X's brain, and (3) this physical continuity has not taken a "branching" form. (4) Personal identity over time just consists in the holding of facts like (2) and (3).

Physical theorists would reject Teletransportation.

The psychological criterion

This involves the continued existence of a purely mental entity, or thing– a soul, or spiritual substance.

The most discussed is the continuity of memory, which can be defined as a branch of psychological criterion. The exceptions are people who have amnesia. Two sets of memories lost due to amnesia: experience memories or fact memories.

Locke suggests that experience-memory provides the criterion of personal identity (not plausible).

Parfit's view:

1. Locke claimed that someone cannot have committed some crimes unless he now remembers doing so. Objection: If true, it would be possible for someone to forget any of the things he once did, or any of the experiences that he once had. But this is possible.
2. If we can remember something happens 20 years ago, they we can say there is a DIRECT MEMORY CONNECTIONS between the two time spots. On Locke's view, only this can make X today and Y 20 years ago the same person. But even if there is no such a connection, namely, if I cannot remember anything 20 years ago, I am still the same person as 20 years ago. Why? Because for each day in 20 years, I can remember, more or less, something happened in the previous day, thus forming an overlapping chain of direct memories.
3. Two general relations: PSYCHOLOGICAL CONNECTEDNESS is the holding of particular direct psychological connections; PSYCHOLOGICAL CONTINUITY is the holding of overlapping chains of strong connectedness, such as which holds between an intention and the later act in which this intention is carried out, or those which hold when a belief, or a desire, or any other psychological feature, continues to be had.
4. On the revised Locke's view, connectedness is more important, since one connection does not merely agrees on personal identity, it

needs thousands of millions of connections. But, connectedness can be hold to any degree.

The psychological criterion: (1) There is psychological continuity if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) it has not amen a "branching" form. (5) Personal identity over time just consists in the holding of facts like (2) to (4).

Three versions of Psy criterion, they differ over the question of what is the right kind of cause.

Narrow version: normal cause.

Wide version: reliable cause.

Widest version: any cause.

e.g. in the teleportation story, my Replica would not be me if we take Physical Criterion and Narrow Psychological Criterion. While if Wide or Widest taken, it would be me.

Parfit: no need to argue over the three subversions, since:

If psychological continuity does not have its normal cause, it may not provide personal identity. But we can claim that, even if this is so, what it provides is as good as personal identity. (blind people, artificial sight, as good as the born one.)

The other views

False assumption about those views above: Materialism or Physicalism–every mental event is just a physical event in some particular brain and nervous system. And other versions, who are not Physicalists are either Dualists or Idealists.

Reductionist: believing that the identity of such a thing may be, in a quite unpuzzling way, indeterminate. Otherwise, non-reductionist.

e.g. The same club? Hard to say. But questioned again for is it the very same club? Absolutely not. The second question is an empty question with only one fact or outcome that we are considering.

When an empty question has no answer, we will give it an answer. So as a Reductionist, sometimes I need to make the same clams, for some empty questions, I need to give myself an answer.

Our identity must be determinate.

Psychological unity is explained by ownership.

What matters in Relation R: psychological connectedness and/or continuity, with the right kind of cause. In an account of what matters, the right kind of cause could be any cause.

How we are not what we believe

Does psychological continuity presuppose personal identity?

No. Define quasi-memory if (1) I seem to remember having an experience, (2) someone did have the experience, and (3) my apparent memory is causally dependent, in the right kind of way, on that past experience.

In this way, the ordinary memories are a subclass of quasi-memories. Quasi implies strong connectedness of quasi-memory. continuity of quasi- memory.

The subject of Experiences

Reid: my personal identity… implies the continued existence of that indivisible thing that I call myself. Whatever this self may be, it is something which thinks, and deliberates, and resolves, and acts, and suffers. I am not thought, I am not action, I am not feeling: I am something that thinks, and acts , and suffers.

Williams's argument against the psychological criterion

Williams's argument seems to refute the Psychological Criterion. It seems to show that the true view is the Physical Criterion. On this view, if some person's brain and body continue to exist, and to support consciousness, this person will continue to exist, however great the breaks are in the psychological continuity of this person's mental life.

The psychological spectrum

Revision of Williams's argument.

A spectrum, or range of cases, each of which is very similar to its neighbours. These cases involve all of the possible degrees of psychological connectedness. I call this the Psychological Spectrum.

The physical spectrum

One objection is that a similar argument applies to physical continuity. Physical Spectrum.

Critical percentage?

The combined spectrum

Physical Spectrum + Psychological Spectrum = Combined Spectrum

The followings are from wiki:

Parfit uses many examples seemingly inspired by Star Trek and other science fiction, such as the teletransporter, to explore our intuitions about our identity. He is a reductionist, believing that since there is no adequate criterion of personal identity, people do not exist apart from their components. Parfit argues that reality can be fully described impersonally; there need not be a determinate answer to the question "Will I continue to exist?" We could know all the facts about a person's continued existence and not be able to say whether the person has survived. He concludes that we are mistaken in assuming that personal identity is what matters; what matters is rather Relation R: psychological connectedness (namely, of memory and character) and continuity (overlapping chains of strong connectedness).

On Parfit's account, individuals are nothing more than brains and bodies, but identity cannot be reduced to either. Parfit concedes that his theories rarely conflict with rival Reductionist theories in everyday life, and that the two are only brought to blows by the introduction of extraordinary examples. However, he defends the use of such examples because they seem to arouse genuine and strong feelings in many of us. Identity is not as determinate as we often suppose it is, but instead such determinacy arises mainly from the way we talk. People exist in the same way that nations or clubs exist.

A key Parfitian question is: given the choice of surviving without psychological continuity and connectedness (Relation R) or dying but preserving R through the future existence of someone else, which would you choose?

Parfit described the loss of the conception of a separate self as liberating:

My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness… [However] When I changed my view, the walls of my glass tunnel disappeared. I now live in the open air. There is still a difference between my life and the lives of other people. But the difference is less. Other people are closer. I am less concerned about the rest of my own life, and more concerned about the lives of others.

Criticism of Personal Identity View:

Fellow reductionist Mark Johnston of Princeton rejects Parfit's constitutive notion of identity with what he calls an "Argument from Above". Johnston maintains, "Even if the lower-level facts [that make up identity] do not in themselves matter, the higher-level fact may matter. If it does, the lower-level facts will have derived significance. They will matter, not in themselves, but because they constitute the higher level fact."

In this, Johnston moves to preserve the significance of personhood. Parfit's explanation is that it is not personhood itself that matters, but rather the facts in which personhood consists that provide it with significance. To illustrate this difference between himself and Johnston, Parfit makes use of an example of a brain-damaged patient who becomes irreversibly unconscious. The patient is certainly still alive even though that fact is separate from the fact that his heart is still beating and other organs are still functioning. But the fact that the patient is alive is not an independent or separately obtaining fact. The patient's being alive, even though irreversibly unconscious, simply consists in the

other facts. Parfit explains that from this so-called “Argument from Below” we can arbitrate the value of the heart and other organs still working without having to assign them derived significance, as Johnston's perspective would dictate.

PART IV: KNOWLEDGE VS. SKEPTICISM

ESSAY 11: Meno, Plato

Sections 96-100

Socrates and Meno (and Anytus, who is largely silent from here on) have now concluded that virtue is at least partly a kind of wisdom, but that even the most beneficent men are not virtuous only out of knowledge (as evidenced by the fact that none of them seem capable of teaching it). This last point, suggests Socrates, is one reason why he and Meno may have failed to find virtue itself in considering such virtuous men. This suggestion puzzles Meno, and Socrates explains that, while they had been looking for virtue as a kind of teachable *knowledge,* virtuous men's good deeds could equally well be the result not of knowledge but of "true opinion."

Socrates gives the example of a guide on the road to Larissa: whether the guide has knowledge of the way or a true opinion about the way, the result is the same (a successful trip to Larissa). But if this is the case, asks Meno, "why is knowledge prized far more highly than right opinion, and why are they different?"

Socrates' answer gives the metaphor of a man who possesses a valuable sculpture by Daedalus. If the statue is "tied down," it is of lasting value. If, however, it is not tied down, it won't last long and is therefore of less good. Similarly, true opinions "are not willing to remain long, and they escape from a man's mind, so that they are not worth much until one ties them down by giving "*an account of the reason why*" the opinion is true [my italics]. Such an account allows true opinion to become knowledge through the process of "recollection" discussed earlier, and so to become fixed in the mind. Nonetheless, at least in terms of directing actions at given times, true opinion serves as well as knowledge.

Socrates and Meno now face a final problem: they have concluded both that virtue cannot be taught and that it is not innate (both parties agree that neither knowledge nor true opinion can be innate). So, returning to the question that opened the dialogue, how do men become virtuous? Plato (through Socrates) is content to leave this a mystery of sorts for now, concluding only that virtuous statesmen are only so

through a sort of divine inspiration, like "soothsayers and prophets. They too say many true things when inspired, but they have no *knowledge* of what they are saying" [my italics].

Thus, virtue is left as "a gift from the gods which is not accompanied by understanding." Though this deep uncertainty may not seem like much of an end to the dialogue, the apparent emptiness of Socrates' conclusion is mitigated by the importance of the lack of knowledge in and of itself. Socrates has succeeded in convincing two prominent citizens and men of politics not only that they have no understanding of virtue, but also that *no one* does. This state of uncertainty, or *aporia*, the state of knowing that one does not know, is a major Platonic theme, and clears the ground for the pursuit of a kind of truth far more exacting and rigorous than had been previously sought.

The *Meno* ends as Socrates bids his interlocutors farewell, reminding them once more that they must seek to know what virtue is (and, according to him, they'd be the first to truly know) before finding out how it comes to be in men. Departing, Socrates tells Meno to teach Anytus what he has learned today.

ESSAY 12: Is Justified True Belief Knowledge? Edmund Gettier

Various attempts have been made in recent years to state necessary and sufficient conditions for someone's knowing a given proposition. The attempts have often been such that they can be stated in a form similar to the following:

a. S knows that P
IFF

1. P is true,
2. S believes that P, and
3. S is justified in believing that P.

For example, Chisholm has held that the following gives the necessary and sufficient conditions for knowledge:

b. S knows that P
IFF

1. S accepts P,
2. S has adequate evidence for P, and
3. P is true.

Ayer has stated the necessary and sufficient conditions for knowledge as follows:

c. S knows that P
IFF

1. P is true,

2. S is sure that P is true, and
3. S has the right to be sure that P is true.

I shall argue that (a) is false in that the conditions stated therein do not constitute a sufficient condition for the truth of the proposition that S knows that P. The same argument will show that (b) and (c) fail if 'has adequate evidence for' or 'has the right to be sure that' is substituted for 'is justified in believing that' throughout.

I shall begin by noting two points. First, in that sense of 'justified' in which S's being justified in believing P is a necessary condition of S's knowing that P, it is possible for a person to be justified in believing a proposition that is in fact false Secondly, for any proposition P, if S is justified in believing P, and P entails Q, and S deduces Q from P and accepts Q as a result of this deduction, then S is justified in believing Q. Keeping these two points in mind, I shall now present two cases in which the conditions stated in (a) are true for some proposition, though it is at the same time false that the person in question knows that proposition.

CASE I

Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:

d. Jones is the man who will get the job, and Jones has ten coins in his pocket.

Smith's evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones's pocket ten minutes ago. Proposition (d) entails:

e. The man who will get the job has ten coins in his pocket.

Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true.

But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e), is false. In our example, then, all of the following are true: (*i*) (e) is true, (*ii*) Smith believes that (e) is true, and (*iii*) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not *know* that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job.

CASE II

Let us suppose that Smith has strong evidence for the following proposition:

f. Jones owns a Ford.

Smith's evidence might be that Jones has at all times in the past within Smith's memory owned a car, and always a Ford, and that Jones has just offered Smith a ride while driving a Ford. Let us imagine, now, that Smith has another friend, Brown, of whose whereabouts he is totally ignorant. Smith selects three place names quite at random and constructs the following three propositions:

g. Either Jones owns a Ford, or Brown is in Boston.

h. Either Jones owns a Ford, or Brown is in Barcelona.

i. Either Jones owns a Ford, or Brown is in Brest-Litovsk.

Each of these propositions is entailed by (f). Imagine that Smith realizes the entailment of each of these propositions he has constructed by (f), and proceeds to accept (g), (h), and (i) on the basis of (f). Smith has correctly inferred (g), (h), and (i) from a proposition for which be has strong evidence. Smith is therefore completely justified in believing each of these three propositions, Smith, of course, has no idea where Brown is.

But imagine now that two further conditions hold. First Jones does *not* own a Ford, but is at present driving a rented car. And secondly, by the sheerest coincidence, and entirely unknown to Smith, the place mentioned in proposition (h) happens really to be the place where Brown is. If these two conditions hold, then Smith does *not* know that (h) is true, even though (*i*) (h) is true, (*ii*) Smith does believe that (h) is true, and (*iii*) Smith is justified in believing that (h) is true.

These two examples show that definition (a) does not state a *sufficient* condition for someone's knowing a given proposition. The same cases, with appropriate changes, will suffice to show that neither definition (b) nor definition (c) do so either.

from wiki:

False premises

n both of Gettier's actual examples, (see also counterfactual conditional), the justified true belief came about, if Smith's purported claims are disputable, as the result of entailment (but see also material conditional) from justified false beliefs that "Jones will get the job" (in case I), and that "Jones owns a Ford" (in case II). This led some early responses to Gettier to conclude that the definition of knowledge could be easily adjusted, so that knowledge was justified true belief that do not depend on false premises.

ESSAY 13: Meditations 1 and 2, Rene Descartes

思想录

First Meditation: skeptical doubts

Summary

The First Meditation, subtitled "What can be called into doubt," opens with the Meditator reflecting on the number of falsehoods he has believed during his life and on the subsequent faultiness of the body of knowledge he has built up from these falsehoods. He has resolved to sweep away all he thinks he knows and to start again from the foundations, building up his knowledge once more on more certain grounds. He has seated himself alone, by the fire, free of all worries so that he can demolish his former opinions with care.

The Meditator reasons that he need only find some reason to doubt his present opinions in order to prompt him to seek sturdier foundations for his knowledge. Rather than doubt every one of his opinions individually, he reasons that he might cast them all into doubt if he can doubt the foundations and basic principles upon which his opinions are founded.

Everything that the Meditator has accepted as most true he has come to learn from or through his senses. He acknowledges that sometimes the senses can deceive, but only with respect to objects that are very small or far away, and that our sensory knowledge on the whole is quite sturdy. The Meditator acknowledges that insane people might be more deceived, but that he is clearly not one of them and needn't worry himself about that.

However, the Meditator realizes that he is often convinced when he is dreaming that he is sensing real objects. He feels certain that he is awake and sitting by the fire, but reflects that often he has dreamed this very sort of thing and been wholly convinced by it. Though his present sensations may be dream images, he suggests that even dream images are drawn from waking experience, much like paintings in that respect. Even when a painter creates an imaginary creature, like a mermaid, the composite parts are drawn from real things--women and fish,

in the case of a mermaid. And even when a painter creates something entirely new, at least the colors in the painting are drawn from real experience. Thus, the Meditator concludes, though he can doubt composite things, he cannot doubt the simple and universal parts from which they are constructed like shape, quantity, size, time, etc. While we can doubt studies based on composite things, like medicine, astronomy, or physics, he concludes that we cannot doubt studies based on simple things, like arithmetic and geometry.

On further reflection, the Meditator realizes that even simple things can be doubted. Omnipotent God could make even our conception of mathematics false. One might argue that God is supremely good and would not lead him to believe falsely all these things. But by this reasoning we should think that God would not deceive him with regard to anything, and yet this is clearly not true. If we suppose there is no God, then there is even greater likelihood of being deceived, since our imperfect senses would not have been created by a perfect being.

The Meditator finds it almost impossible to keep his habitual opinions and assumptions out of his head, try as he might. He resolves to pretend that these opinions are totally false and imaginary in order to counter-balance his habitual way of thinking. He supposes that not God, but some evil demon has committed itself to deceiving him so that everything he thinks he knows is false. By doubting everything, he can at least be sure not to be misled into falsehood by this demon.

Analysis

The First Meditation is usually approached in one of two ways. First, it can be read as setting the groundwork for the meditations that follow, where doubt is employed as a powerful tool against Aristotelian philosophy. Second, it can, and often is, read standing on its own as the foundation of modern skepticism. We will briefly discuss these complementary readings in turn.

Descartes saw his *Meditations* as providing the metaphysical underpinning of his new physics. Like Galileo, he sought to overturn two-thousand-year-old prejudices injected into the Western tradition by Aristotle. The Aristotelian thought of Descartes' day placed a great weight on the testimony of the senses, suggesting that all knowledge comes from the senses. The Meditator's suggestion that all one's most certain knowledge comes from the senses is meant to appeal directly to the Aristotelian philosophers who will be reading the *Meditations*. The motivation, then, behind the First Meditation is to start in a position the Aristotelian philosophers would agree with and then, subtly, to seduce them away from it. Descartes is aware of how revolutionary his ideas are, and must pay lip service to the orthodox opinions of the day in order to be heeded.

Reading the First Meditation as an effort to coax Aristotelians away from their customary opinions allows us to read different interpretations into the different stages of doubt. For instance, there is some debate as to whether Descartes intended his famous "Dream Argument" to suggest the universal possibility of dreaming--that though there is waking experience, I can never know which moments are dreams and which are waking--or the possibility of a universal dream--that my whole life is a dream and that there is no waking world. If we read Descartes as suggesting the universal possibility of dreaming, we can explain an important distinction between the Dream Argument and the

later "Evil Demon Argument." The latter suggests that all we know is false and that we cannot trust the senses one bit. The Dream Argument, if meant to suggest the universal possibility of dreaming, suggests only that the senses are not always and wholly reliable. The Dream Argument questions Aristotelian epistemology, while the Evil Demon Argument does away with it altogether. The "Painter's Analogy," which draws on the Dream Argument, concludes that mathematics and other purely cerebral studies are far more certain than astronomy or physics, which is an important step away from the Aristotelian reliance on the senses and toward Cartesian rationalism.

The *Meditations* can be seen to follow the model of St. Ignatius of Loyola's *Spiritual Exercises*. The first step in the Jesuit exercises is to purge oneself of one's attachment to the material, sinful world. In the First Meditation, Descartes leads us through a similar purgation, though with a different purpose. Here he wants to persuade his Aristotelian readers to purge themselves of their prejudices. He also hopes to lead the mind away from the senses that are so heavily relied upon by the Aristotelians. In the meditations that follow, he will argue that our most certain knowledge comes from the mind unaided by the senses. Lastly, this process of radical doubt will hopefully rule out any doubts from the positive claims Descartes will build up in the next five meditations. Read in the wider context of the *Meditations,* these skeptical doubts are a means to the end of preparing a resistant audience to the metaphysics Descartes plans to build.

Read on its own, the First Meditation can be seen as presenting skeptical doubts as a subject of study in their own right. Certainly, skepticism is a much discussed and hotly debated topic in philosophy, even today. Descartes was the first to raise the mystifying question of how we can claim to know with certainty anything about the world around us. The idea is not that these doubts are probable, but that their possibility can never be entirely ruled out. And if we can never be certain, how can we claim to know anything? Skepticism cuts straight to the heart of the Western philosophical enterprise and its attempt to provide a certain foundation for our knowledge and understanding of the world. It can even be pushed so far as to be read as a challenge to our very notion of rationality.

No one actually lives skepticism--no one actually doubts whether other people really exist--but it is very difficult to justify a dismissal of skepticism. Western philosophy since Descartes has been largely marked and motivated by an effort to overcome this problem. Particularly interesting responses can be found in Hume, Kant, and Wittgenstein.

We should note that Descartes' doubt is a methodological and rational doubt. That is, the Meditator is not just doubting everything at random, but is providing solid reasons for his doubt at each stage. For instance, he rejects the possibility that he might be mad, since that would undercut the rationality that motivates his doubt. Descartes is trying to set up this doubt within a rational framework, and needs to maintain a claim to rationality for his arguments to proceed.

Second Meditation, Part 1: cogito ergo sum and sum res cogitans

Summary

The Second Meditation is subtitled "The nature of the human mind, and how it is better known than the body" and takes place the day after the First Meditation. The Meditator is firm in his resolve to continue his search for certainty and to discard as false anything that is open to the slightest doubt. He recalls Archimedes' famous saying that he could shift the entire earth given one immovable point: similarly, he hopes to achieve great things if he can be certain of just one thing. Recalling the previous meditation, he supposes that what he sees does not exist, that his memory is faulty, that he has no senses and no body, that extension, movement and place are mistaken notions. Perhaps, he remarks, the only certain thing remaining is that there is no certainty.

Then, he wonders, is not he, the source of these meditations, not something? He has conceded that he has no senses and no body, but does that mean he cannot exist either? He has also noted that the physical world does not exist, which might also seem to imply his nonexistence. And yet to have these doubts, he must exist. For an evil demon to mislead him in all these insidious ways, he must exist in order to be misled. There must be an "I" that can doubt, be deceived, and so on. He formulates the famous *cogito* argument, saying: "So after considering everything very thoroughly, I must finally conclude that this proposition, *I am, I exist,* is necessarily true whenever it is put forward by me or conceived in my mind."

The Meditator's next question, then, is what this "I" that exists is. He initially thought that he had a soul, by means of which he was nourished, moved, could sense and think; and also that he had a body. All these attributes have been cast into doubt, except one: he cannot doubt that he thinks. He may exist without any other of the above attributes, but he cannot exist if he does not think. Further, he only exists as long as he is thinking. Therefore, thought above all else is inseparable from being. The Meditator concludes that, in the strict sense, he is only a thing that thinks.

Analysis

The *cogito* argument is so called because of its Latin formulation in the *Discourse on Method*: "*cogito ergo sum*" ("I think, therefore I am"). This is possibly the most famous single line in all of philosophy, and is generally considered the starting point for modern Western philosophy. In it, the Meditator finds his first grip on certainty after the radical skepticism he posited in the First Meditation. The *cogito* presents a picture of the world and of knowledge in which the mind is something that can know itself better than it can know anything else. The idea that we know our

mind first and foremost has had a hypnotic hold on Western philosophy ever since, and how the mind can connect with reality has ever since been a major concern. In this conception, the mind ceases to be something that helps us know about the world and becomes something inside which we are locked.

We should note, however, the distinction between the "I think, therefore I am" as stated in the *Discourse on Method* and the formulation we get in the *Meditations*: "So after considering everything very thoroughly, I must finally conclude that this proposition, *I am, I exist,* is necessarily true whenever it is put forward by me or conceived in my mind." Neither "therefore" nor "I think" appear in the *Meditations.* The absence of "therefore" is important, since it dissuades us from reading the *cogito* as a syllogism, that is, as a three-step argument as follows:

(1) Whatever thinks exists

(2) I think

Therefore (3) I exist

The trouble with a syllogistic reading, which Descartes explicitly denies elsewhere in his writings, is that no reason is given why (1) should be immune from the doubt that the Meditator has posited. Also, the syllogistic reading interprets the *cogito* as a reasoned inference at a point in the Meditator's doubt when even reasoned inferences can be called into doubt.

But if everything is to be doubted, how can the Meditator know the *cogito*? A number of readings have been given to understand this step. One is to read it as an intuition rather than an inference, as something that comes all at once, in a flash. Another reading interprets the *cogito* as a performative utterance, where the utterance itself is what confirms its truth. That is, I could not say "I exist" if I did not exist or if I did not think, and so the act of saying it is what makes it true. Thus, I can only affirm my own existence (not anybody else's) and I can only do so in the present tense: I cannot say "I thought, therefore I was/am."

It should be noted that the *cogito* only works for thought. I cannot say, "I walk, therefore I am," since I can doubt I am walking. The reason I cannot doubt that I am thinking is that doubt itself is a form of thought.

After the *cogito,* the Meditator advances the claim that he is a thing that thinks, an argument called the *sum res cogitans,* after its Latin formulation. There are three controversies regarding the claim "I am...in the strict sense only a thing that thinks," which we will examine in turn: whether the claim is metaphysical or epistemological, what is meant by "thing," and what is meant by "thinking."

It is more plausible to read the *sum res cogitans* as an epistemological remark, saying that, "whatever else I may be, I know only that I am a thing that thinks." However, in some of his writings, Descartes makes it plausible to read him as making a metaphysical remark, that "I *am only* a thing that thinks." His reasoning might go something like this: "I know that I am a thinking thing, and I do not know whether I am a bodily thing. My body and my mind cannot be one and the same, because I should either know both of them or know neither of them. Since I know I am a thinking thing, and know that my body and my mind are two separate things, I can conclude that I am not a bodily thing. There-

fore, I am only a thing that thinks." In so arguing, however, Descartes would commit the so-called "intentional fallacy" of basing an argument on what one does not know. If two things had to be either both known or both not known in order to be identical, we could argue that Bruce Wayne and Batman are not one and the same as well.

"Thing that thinks" also carries some ambiguous baggage. By "thing," Descartes could simply be using the word as we do today, as an ambiguous throwaway word when we don't want to be more specific. More likely, though, he is using it to mean substance, the fundamental and indivisible elements of Cartesian ontology. In this ontology, there are extended things (bodies) and thinking things (minds), and Descartes is here asserting that we are minds rather than bodies. Of course, "thinking" is also highly questionable. Does Descartes mean only the intellection and understanding that is characteristic of the Aristotelian conception of mind? Or does he also include sensory perception, imagination, willing, and so on? At the beginning of the Second Meditation, the Meditator has cast sensory perception and so on into doubt, but by the end of the Second Meditation, sensing, imagining, willing, and so on are included as attributes of the mind. This question is further explored in the commentary on the next section.

Second Meditation, Part 2: the wax argument

Summary

The Meditator tries to clarify precisely what this "I" is, this "thing that thinks." He concludes that he is not only something that thinks, understands, and wills, but is also something that imagines and senses. After all, he may be dreaming or deceived by an evil demon, but he can still imagine things and he still *seems* to hear and see things. His sensory perceptions may not be veridical, but they are certainly a part of the same mind that thinks.

The Meditator then moves on to ask how he comes to know of this "I." The senses, as we have seen, cannot be trusted. Similarly, he concludes, he cannot trust the imagination. The imagination can conjure up ideas of all sorts of things that are not real, so it cannot be the guide to knowing his own essence. Still, the Meditator remains puzzled. If, as he has concluded, he is a thinking thing, why is it that he has such a distinct grasp of what his body is and has such a difficult time identifying what is this "I" that thinks? In order to understand this difficulty he considers how we come to know of a piece of wax just taken from a honeycomb: through the senses or by some other means?

He first considers what he can know about the piece of wax by means of the senses: its taste, smell, color, shape, size, hardness, etc. The Meditator then asks what happens when the piece of wax is placed near the fire and melted. All of these sensible qualities change, so that, for

instance, it is now soft when before it was hard. Nonetheless, the same piece of wax still remains. Our knowledge that the solid piece of wax and the melted piece of wax are the same cannot come through the senses since all of its sensible properties have changed.

The Meditator considers what he can know about the piece of wax, and concludes that he can know only that it is extended, flexible, and changeable. He does not come to know this through the senses, and realizes that it is impossible that he comes to know the wax by means of the imagination: the wax can change into an infinite number of different shapes and he cannot run through all these shapes in his imagination. Instead, he concludes, he knows the wax by means of the intellect alone. His mental perception of it can either be imperfect and confused--as when he allowed herself to be led by his senses and imagination-- or it can be clear and distinct--as it is when he applies only careful mental scrutiny to his perception of it.

The Meditator reflects on how easy it is to be deceived regarding these matters. After all, we might say "I see the wax," though in saying that we refer to the wax as the intellect perceives it, rather than to its color or shape. This is similar to the way in which we might "see" people down in the street when all we really see are coats and hats. Our intellect--and not our eyes--judges that there are people, and not automata, under those coats and hats.

The Meditator concludes that, contrary to his initial impulses, the mind is a far better knower than the body. Further, he suggests, he must know his mind far better than other things. After all, as he has admitted, he may not be perceiving the piece of wax at all: it may be a dream or an illusion. But when he is perceiving the piece of wax, he cannot doubt that he is perceiving nor that he is judging what he perceives to be a piece of wax, and both of these acts of thought imply that he exists. Every thought we might have about the world outside us can only doubtfully be true of the outside world, but it must with certainty confirm our own existence and establish the nature of our own mind.

The Meditator happily concludes that he can know at least that he exists, that he is a thinking thing, that his mind is better known than his body, and that all clear and distinct perceptions come by means of the intellect alone, and not the senses or the imagination.

Analysis

The first paragraph of the above summary covers the ninth paragraph of the Second Meditation. We could identify this moment as the invention of the modern mind. The Aristotelian conception of the mind separates intellection and understanding as attributes of a soul that survive death. Sensing, imagining, willing, etc., are all attached to the sensory world and are therefore distinct, according to Aristotle. In the Cartesian conception of mind, there is a sharp distinction between mind and world, where all those activities--like sensing and imagining--that could take place in dreams or in disembodied minds are considered mental activities, and exist only in the mind. Things in the world such as trees or light waves are then totally separate from things in the mind, and it becomes a major concern for modern philosophy to determine how the two connect. For instance, there seems to be some connection between my visual sensations and the objects in the world that I see, but since visual sensations are a part of the mind and the objects I see are a part of the world, it is very difficult to determine what that con-

nection is. This picture of mind may seem intuitive to us now, but it and the theories of mind that have sprung from it originate in Descartes. Only in the twentieth century have philosophers like Wittgenstein, William James, and J. L. Austin come to question Descartes' sharp distinction between mind and world.

The rest of the Second Meditation concentrates on the "Wax Argument" with which Descartes hopes to show definitively that we come to know things through the intellect rather than through the senses and that we know the mind better than anything else. His argument focuses on the process of change by which solid wax melts into a liquid puddle. The senses seem to tell us things about the world, and Descartes admits that what we know about the solid piece of wax we know through the senses. The senses can similarly inform us about the melted wax, but they cannot tell us that the melted wax and the solid wax are one and the same. Nor, Descartes argues, can the imagination. Only the intellect can organize and make sense of what we perceive. The senses only perceive a disconnected jumble of information: the intellect is what helps us to understand it.

This argument is another move against the Aristotelian theory of knowledge, according to which all knowledge comes from the senses. Descartes acknowledges that the senses inform us about the world, but asserts that the senses can only give us disorganized information. Without the intellect, we could make no sense of what we perceive. Descartes thus places himself firmly in the rationalist camp, as compared to empiricists such as Aristotle or Locke who argue for a sense-based theory of knowledge.

Descartes' next move is a little more questionable. He asserts that "I" cannot know with certainty that what "I" perceive is real (as per the doubts of the First Meditation), but that sensory perception, as a form of thought, confirms that "I" exist ("I" being the mind.) Every time "I" perceive "I" am thinking, and in thinking "I" am enacting the *cogito.* Every perception confirms the existence of "my" mind and only gives dubitable evidence for the existence of the world. Thus, Descartes concludes, the mind is better known than the body.

This argument is plausible if Descartes means that the existence of the mind is better known than the existence of the body, but it seems that he wants to say that the nature of the mind is better known than the nature of the body. That is, Descartes wants to say that "I" know not only that the mind exists, but also "I" know more about the mind than about the world outside the mind. This argument would only hold if every thought, perception, imagination, etc., told "me" something new about the mind. But, according to the *cogito,* all these thoughts tell "me" only one and the same thing: that "I" exist, and that "I" am a thing that thinks. Descartes is not as clear as we might like him to be as to what and how exactly each new thought makes the mind better known than the body.

ESSAY 14: Proof of an External World, G. E. Moore

One of the most important parts of Moore's philosophical development was his break from the idealism that dominated British philosophy (as represented in the works of his former teachers F. H. Bradley and John McTaggart), and his defence of what he regarded as a "common sense" form of realism. In his 1925 essay "A Defence of Common Sense", he argued against idealism and scepticism toward the external world on the grounds that they could not give reasons to accept their metaphysical premises that were more plausible than the reasons we have to accept the common sense claims about our knowledge of the world that sceptics and idealists must deny. He famously put the point into dramatic relief with his 1939 essay "Proof of an External World", in which he gave a common sense argument against scepticism by raising his right hand and saying "Here is one hand," and then raising his left and saying "And here is another," then concluding that there are at least two external objects in the world, and therefore that he knows (by this argument) that an external world exists. Not surprisingly, not everyone inclined to sceptical doubts found Moore's method of argument entirely convincing; Moore, however, defends his argument on the grounds that sceptical arguments seem invariably to require an appeal to "philosophical intuitions" that we have considerably less reason to accept than we have for the common sense claims that they supposedly refute. (In addition to fueling Moore's own work, the "Here is one hand" argument also deeply influenced Wittgenstein, who spent his last years working out a new approach to Moore's argument in the remarks that were published posthumously as *On Certainty*.)

Here is one hand

Here is one hand is the name of a philosophical argument created by George Edward Moore against philosophical skepticism and in support of common sense.

The argument takes the form:

- Here is one hand,

- And here is another.
- There are at least two external objects in the world.
- Therefore an external world exists.

G. E. Moore (1873—1958) wrote *A Defence of Common Sense* and *Proof of an External World*. He posed skeptical hypotheses, such as "you may be dreaming" or "the world is 5 minutes old", creating a situation where it is not possible to know that anything in the world exists. These hypotheses take the following form:

The skeptical argument

Where **S** is a subject, **sp** is a skeptical possibility, such as the brain in a vat hypothesis, and **q** is a knowledge claim about the world:

- If **S** doesn't know that not-**sp**, then **S** doesn't know that **q**
- **S** doesn't know that not-**sp**
- Therefore, **S** doesn't know that **q**

Moore's response

- If **S** doesn't know that not-**sp**, then **S** doesn't know that **q**
- **S** knows that **q**
- Therefore, **S** knows that not-**sp**

Moore does not attack the skeptical premise; instead, he reverses the argument from being in the form of modus ponens to modus tollens. This logical maneuver is often called a **G. E. Moore shift** or a **Moorean shift.**

Explanation

Moore famously put the point into dramatic relief with his 1939 essay *Proof of an External World*, in which he gave a common sense argument against skepticism by raising his right hand and saying "here is one hand," and then raising his left and saying "and here is another". Here, Moore is taking his knowledge claim (**q**) to be that he has two hands, and without rejecting the skeptic's premise, proves that we can know the skeptical possibility (**sp**) to be not true.

Moore's argument is not simply a flippant response to the skeptic. Moore gives in *Proof of an External World*, three requirements for a good proof. (1) the premises must be different from the conclusion, (2) the premises must be demonstrated, and (3) the conclusion must follow from the premises. He claims that his proof of an external world meets those three criteria.

In his 1925 essay *A Defence of Common Sense* he argued against idealism and skepticism toward the external world on the grounds that they

could not give reasons to accept their metaphysical premises that were more plausible than the reasons we have to accept the common sense claims about our knowledge of the world that skeptics and idealists must deny. In other words, he is more willing to believe that he has a hand than believe the premises of a strange argument in a university classroom. "I do not think it is rational to be as certain of any one of these ... propositions".

Not surprisingly, those inclined to skeptical doubts often found Moore's method of argument not entirely convincing. Moore, however, defends his argument on the surprisingly simple grounds that skeptical arguments seem invariably to require an appeal to "philosophical intuitions" that we have considerably less reason to accept than we have for the common sense claims that they supposedly refute.

Logical form

The skeptical argument takes the form of *modus ponens*:

- If **A** then **B**.
- **A**.
- Therefore **B**.

Moore's argument flips the modus ponens structure into a *modus tollens*:

- If **A** then **B**.
- **Not B**.
- Therefore **not A**.

This illustrates Fred Dretske's aphorism that "[o]ne man's *modus ponens* is another man's *modus tollens*"

ESSAY 15: Certainty, G. E. Moore

Moore is concerned with the following argument:

P1) If it is not certain that I am not dreaming, then it is not certain that I am standing up.
P2) I am not certain that I am not dreaming.
C) I am not certain that I am standing up.

This is, of course, an argument that is familiar from Descartes and others. But is it a good argument?

P1) If I don't know that I am not dreaming, then I don't know that I am standing up.
P2) I don't know that I am not dreaming.
C) I don't know that I am standing up.

Moore thinks that P1 is true. Why?

Suppose I dream that I am sitting in a chair in Moscow. As it turns out, I am sitting in a chair in Moscow—this is where I've fallen asleep. While dreaming, I've got a belief (that I'm sitting in a chair in Moscow) that turns out to be true. Do we want to say that I know I'm sitting in a chair in Moscow?

P1) If I don't know that I am not dreaming, then I don't know that I am standing up.
P2) I don't know that I am not dreaming.

C) I don't know that I am standing up.
No. Knowledge requires more than this. To say that I know some proposition P, the following conditions need to be met:
1) I must believe that P.
2) I must be justified in believing that P.
3) P must be true.
On this view, knowledge is understood as justified true belief. We'll call this the JTB theory of knowledge.
P1) If I don't know that I am not dreaming, then I don't know that I am standing up.
P2) I don't know that I am not dreaming.
C) I don't know that I am standing up.
So P1 looks safe to Moore. If I don't know that I am not dreaming, then my belief that I am standing up is not justified.
This is not enough, though, to persuade Moore that he should accept C.
"This first part of the argument is a consideration which cuts both ways":
P1) If I don't know that I am not dreaming, then I don't know that I am standing up.
P2) I know that I am standing up.
C) I know that I am not dreaming.
Notice that this argument is formally valid. As Moore says, "The one argument is just as good as the other, unless my opponent can give better reasons for asserting that I don't know that I'm not dreaming, than I can give for asserting that I do know that I am standing up."
Here's a similar argument against

skepticism about the external world:
P1) If I am unsure that the external world exists, then I am unsure that I have hands.
P2) I am sure that I have hands.
C) I am sure that the external world exists.

ESSAY 16: The Matrix as Metaphysics, David J. Chalmers

Hypothesis (Matrix)

I am in a matrix; or, equivalently, I am envatted and have always been envatted.

Matrix Hypothesis seems to be a **skeptical hypothesis** in that (if true) it seems to render almost all of my beliefs concerning the world false

Reasoning: "I don't know that I'm not in a matrix. If I'm in a matrix, I'm probably not in Tucson. The same goes for almost everything else I think I know about the external world." (135)

Chalmers:

I cannot rule out the Matrix Hypothesis.

But: it is not a skeptical hypothesis **because even if it's true, most of my beliefs will still be true!**

For instance: even if I'm envatted, I am still walking outside in the sun in Tucson

Instead, it's a **metaphysical hypothesis** , i.e. it concerns the

nature of the most fundamental level of reality.

Chalmers: in fact, the Matrix Hypothesis is equivalent to the

conjunction of three hypotheses:

1 Computational Hypothesis

2 Creation Hypothesis

3 Mind-Body Hypothesis

The Computational Hypothesis

"Microphysical processes throughout space-time are constituted by underlying computational processes."

don't know that it is true, don't know that it is false, but it's coherent

not skeptical: elementary particles etc are just more like tables and chairs, and so fundamental reality is different from what we thought, but it still exists and most of our ordinary beliefs are not affected by its truth or falsity

The Creation Hypothesis

"Physical space-time and its contents were created by beings outside physical space-time."

don't know that it is true, don't know that it is false, but it's coherent

not skeptical: even if true, most of my ordinary beliefs remain valid

The Mind-Body Hypothesis

"My mind is (and has always been) constituted by processes outside physical space-time and receives its perceptual inputs from and sends its outputs to processes in physical space-time."

don't know that it is true, don't know that it is false, but it's coherent

not skeptical: even if true, most of my ordinary beliefs remain valid

The Metaphysical Hypothesis

Physical space-time and its contents were created by beings outside physical space-time. Both physical space-time and the microphysical processes it contains are constituted by computational processes that were designed as a computer simulation of the world. Also, our minds are outside physical space-time but interact with it.

The Matrix Hypothesis as a Metaphysical Hypothesis

Again, Metaphysical Hypothesis is **coherent** and **not skeptical**.

Chalmers: Matrix Hypothesis is **equivalent** to Metaphysical

Hypothesis, i.e. they imply one another

Metaphysical -> Matrix: from Mind-Body, Computational and

Creation Hypotheses, it follows that "I have (and always had) a

cognitive system that receives its input from and sends its output to an artifically designed computer simulation of the world", but that's just the Matrix Hypothesis

Matrix -> Metaphysical: accepting Matrix means to accept that

whatever underlies apparent reality is really just as Metaphysical Hyp claims, viz. that there is a domain containing my mind, which causally interacts with an artificially created computer simulation of physical space-time and its contents

Conclusion: anti-skepticism

If this is right, then the Matrix Hypothesis is **not a skeptical hypothesis** as the resulting picture is one of a "full-blooded external world",

even though it entails that fundamental reality is quite a bit different from what our currently best scientific theories tell us.

Let's consider a few objections.

1. Envatted brain may think it is in Tucson when in fact it is in Sydney. Response: the envatted brain's concept of "Tucson" does not refer to Tucson, but to something else entirely (call it "Tucson*")*
2. But what sort of thing does the envatted being refer to?Response: entities constituted by computational processes

A truly skeptical hypothesis

Hypothesis (Chaos)

"I do not receive inputs from anywhere in the world. Instead, I have random, uncaused experience. Through a huge coincidence, they are exactly the sort of regular, structured experiences with which I am familiar."

coherent, but has minuscule probability

truly **skeptical**: if true, almost all of our beliefs would be true, and almost none of our concepts could refer (to physical objects, or patterns of bits in computational processes)

The skeptical argument

Brain floating in nutrient fluids is disembodied. ("The Matrix" depicts embodied brains).

Important: the brain has experiences which are qualitatively

indistinguishable from those of normal perceiver.

Skeptical challenge reformulated: on what grounds can you rule

out this possibility? Skeptic argues:

If you know that **p**, then you know that you are not a brain in a vat.

You don't know that you are not a brain in a vat.

So, you don't know that **p**.

PART V: ETHICS

ESSAY 17: Glaucon's Challenge, Plato

The Republic, Book II

Summary: Book II, 357a–368c

Socrates believes he has adequately responded to Thrasymachus and is through with the discussion of justice, but the others are not satisfied with the conclusion they have reached. Glaucon, one of Socrates's young companions, explains what they would like him to do. Glaucon states that all goods can be divided into three classes: things that we desire only for their consequences, such as physical training and medical treatment; things that we desire only for their own sake, such as joy; and, the highest class, things we desire both for their own sake and for what we get from them, such as knowledge, sight, and health. What Glaucon and the rest would like Socrates to prove is that justice is not only desirable, but that it belongs to the highest class of desirable things: those desired both for their own sake and their consequences.

Glaucon points out that most people class justice among the first group. They view justice as a necessary evil, which we allow ourselves to suffer in order to avoid the greater evil that would befall us if we did away with it. Justice stems from human weakness and vulnerability. Since we can all suffer from each other's injustices, we make a social contract agreeing to be just to one another. We only suffer under the burden of justice because we know we would suffer worse without it. Justice is not something practiced for its own sake but something one engages in out of fear and weakness.

To emphasize his point, Glaucon appeals to a thought experiment. Invoking the legend of the ring of Gyges, he asks us to imagine that a just man is given a ring which makes him invisible. Once in possession of this ring, the man can act unjustly with no fear of reprisal. No one can deny, Glaucon claims, that even the most just man would behave unjustly if he had this ring. He would indulge all of his materialistic, power-hungry, and erotically lustful urges. This tale proves that people are only just because they are afraid of punishment for injustice. No one is just because justice is desirable in itself.

Glaucon ends his speech with an attempt to demonstrate that not only do people prefer to be unjust rather than just, but that it is rational for them to do so. The perfectly unjust life, he argues, is more pleasant than the perfectly just life. In making this claim, he draws two detailed portraits of the just and unjust man. The completely unjust man, who indulges all his urges, is honored and rewarded with wealth. The completely just man, on the other hand, is scorned and wretched.

His brother, Adeimantus, breaks in and bolsters Glaucon's arguments by claiming that no one praises justice for its own sake, but only for the rewards it allows you to reap in both this life and the afterlife. He reiterates Glaucon's request that Socrates show justice to be desirable in the absence of any external rewards: that justice is desirable for its own sake, like joy, health, and knowledge.

Analysis: Book II, 357a–368c

Coming on the heels of Thrasymachus' attack on justice in Book I, the points that Glaucon and Adeimantus raise—the social contract theory of justice and the idea of justice as a currency that buys rewards in the afterlife—bolster the challenge faced by Socrates to prove justice's worth. With several ideas of justice already discredited, why does Plato further complicate the problem before Socrates has the chance to outline his own ideas about justice?

The first reason is methodological: it is always best to make sure that the position you are attacking is the strongest one available to your opponent. Plato does not want the immoralist to be able to come back and say, "but justice is only a social contract" after he has carefully taken apart the claim that it is the advantage of the stronger. He wants to make sure that in defending justice, he dismantles all the best arguments of the immoralists.

The accumulation of further ideas about justice might be intended to demonstrate his new approach to philosophy. In the early dialogues, Socrates often argues with Sophists, but Thrasymachus is the last Sophist we ever see Socrates arguing with. From now on, we never see Socrates arguing with people who have profoundly wrong values. There is a departure from the techniques of *elenchus* and *aporia*, toward more constructive efforts at building up theory.

The Republic was written in a transitional phase in Plato's own life. He had just founded the Academy, his school where those interested in learning could retreat from public life and immerse themselves in the study of philosophy. In his life, Plato was abandoning Socrates's ideal of questioning every man in the street, and in his writing, he was abandoning the Sophist interlocutor and moving toward conversational partners who, like Glaucon and Adeimantus, are carefully chosen and prepared. In the dialogues, they are usually Socrates's own students.

Plato had decided at this point that philosophy can only proceed if it becomes a cooperative and constructive endeavor. That is why in his own life he founded the Academy and his writings paired Socrates with partners of like mind, eager to learn. Glaucon and Adeimantus repeat the challenge because they are taking over the mantle as conversational partners. Discussion with the Sophist Thrasymachus can only lead to *aporia*. But conversation with Glaucon and Adeimantus has the potential to lead to positive conclusions.

This might seem like a betrayal of his teacher's mission, but Plato probably had good reason for this radical shift. Confronting enemies has severe limits. If your viewpoint differs radically from that of your conversational partner, no real progress is possible. At most, you can undermine one another's views, but you can never build up a positive theory together.

ESSAY 18: The Challenge of Cultural Relativism, James Rachels

See class notes

ESSAY 19: Egoism and Moral Skepticism, James Rachels

Descriptive: describe ethical/moral behavior

Normative: describe how we OUGHT to act

Hume's Fork: There's no theoretical way to get from an "Is" to an "Ought"

"Ought" implies "Can" : if you should do something (x), that means it's possible for that thing to be done.

Psychological Egoism

Descriptive: All actions ARE self-interested; every action we do is motivated by self-interest

Two Arguments in Favor Of it

- A) Regardless of what we say our reasons for action are, the fact of the matter is that we always do what we want to do. (Smith eg on p 13)
- Rachels REJECTS This 2 Ways
  - I) Assumes that people only do what they want to do. There at least 2 cases where people do things they don't want to do.
    - We do undesirable actions all the time in order to achieve desirable ends. (Like the dentist)
    - When we feel we have made obligations to do something (Keeping a promise).
  - 2) Acting on my desires doesn't necessarily mean that the action is selfish. We need to look at the object of our desires. Acting exclusively on my interests is selfish, but acting in a way (that I want to act) that also promotes the interest of others is not selfish. (Intention)
- B) Unselfish actions only superficially appear to be unselfish. We really do unselfish action because of the great feelings we get from doing such things. Therefore, such actions are really selfish. (Lincoln example 14)

- Rachels REJECTS this:
  - Feeling good feelings from doing action doesn't mean that you do the action for selfish reasons. Instead, it means that you're the kind of person who is motivated by helping others.

If Psychological Egoism is, as Rachels's has shown, obviously so bad, then why do people think it's a legitimate view?

Because people get confused about three sets of distinctions:

- 1) Selfishness v. Self-interest
  - Self-interest : acting in your own interests.
  - Selfishness: Acting on your interests in such a way that you ignore the interests of others in circumstance where their interests should NOT be ignored
- 2) Action are done EITHER from S-interest OR Other-regarding interests (Altruism) (False Dichotomy)
  - Not True: We do lots of things that are neither in our interest nor in the interest of others (Smoking)
    - (Not all actions are selfish and not all actions are done from self-interest. Rachels has done this without appeal to altruism)
- 3) Concern for you well-being is incompatible with concern for the well-being of others
  - Also False: We can promote the interests of others and our own interests at the same time. There's no necessary inconsistency there.

Ethical Egoism

Normative: All actions OUGHT to be self-interested. Regardless of how we do act, we are under no obligation to act in any way except in ways that promote our own interests.

Rachels s hows a problem with what Ethical Egoism may require :

- Does this mean that I could (that I SHOULD) burn down a building if I got pleasure fromdoing it? (16)
  - The "standard" ethical egoist response:
    - NO! You need to have more foresight. What's really in your interest is to have a society that's law abiding. This will ensure your protection, and that's integral to your interests.
      - It's to your own advantage to be "good" since that will serve your interests in the long run.

- Rachels Refutes the Response
  - Well… Obviously we want a stable society, but a few law breakers won't undermine the social fabric.
  - What the ethical egoist really needs to do is convince others to act for long term stability, while at the same time the ethical egoist acts on his or her interests.
  - The ethical egoist pretends to be a do-gooder, while really being completely self-interested.
    - This is kind of the point of the Ring of Gyges story.
- Problem of Universality
  - Generally, all ethical theories should apply equally to everyone. If each egoist wants to public endorse altruism while privately endorsing egoism, it seems like ethical egoism fails the universality requirement.
  - Rachels offers a solution to this.
    - There's no logical inconsistency if we frame the egoist's ideal world as one where his or her interests are promoted by everyone.
    - There is a very high chance that this is practically impossible.
- Interesting disucssion on page 18 about Harm.

In the End: Pyschological Egoism is blatantly False. Ethical Egoism has some problems, but seems to at least be plausible.

ESSAY 20: What Makes Someone's Life Go Best, Derek Parfit

Answer to this is called *theories about self-interest*, three kinds:

1. *Hedonistic Theories*: what would be best for someone is what would make his life happiest.
2. *Desire-Fulfilment Theories*: ...is what, throughout his life, would best fulfil his desires.
3. *Objective List Theories*: certain things are good or bad for us, whether or not we want to have the good things, or to avoid the bad things.

1.1 *Narrow Hedonists*: assume, falsely, that pleasure and pain are two distinctive kinds of experience, they basically do not have common quality.

1.2 *Preference-Hedonism*: what they have in common are their relations to our desires. Pains are unwanted experiences, pleasures are wanted ones. e.g. reading rather than going to party

2.1 *Unrestricted D-F Theory*: what is best for someone is what would best fulfil all of his desires, throughout his life. e.g. my sympathy for strangers with fatal disease, then he is cued, so it's good for me, my life goes better. (reject this theory)

2.2 *Success Theory*: Only one thing differs from Preference-Hedonism, it appeals to all of our preferences about our own lives. e.g. if my belief is false though, then Preference-Hedonism would consider it as a preference, while Success Theory would consider it as wrong.

Cambridge-change: An object undergoes a Cambridge-change if there is any change in the true statements that can be made about this object. Suppose that I cut my cheek while shaving. This causes a real change in me. It also causes a change in Confucius. It becomes true, of Confucius, that he lived on a planet on which later one more cheek was cut. This is merely a Cambridge-change.

about ex: an exile parent does not know that his children live badly, he makes some choices for them and wants to be a good parent. From P-H, he is a good parent, but from Success Theory, he fails to be such one.

Should we appeal only to the desires and preferences that someone actually has?

Yes, although I chose to read King Lear, rather than going to a party, I may think opposite if I really went there. **each alternative would have been better than the other.**

Whether we appeal to Preference-Hedonism or the Success Theory, we should not appeal only to the desires or preferences that I actually have. WE should also appeal to the desires and preferences that I would have had, in the various alternatives that were, at different times, open to me.

Another distinction for both: *Summative*. The total net sum of desire-fulfilment is the sum of the positive numbers minus the negative numbers.

Another version of both theories does not appeal to all of a person's desires and preferences about his own life. global > local. A preference is global if it is about some part of one's whole life. The global versions are more plausible.

Case of addiction:

Summative Theories: making you an addict => increasing the sum-total of your desire-fulfilment => causing one desire (not to be an addict) not fulfilled => replaced by anther desire (to be cued) => but overweighs, so i am benefiting your life.

> *This imagined case of addiction is in its essentials similar to countless other cases. There are countless cases in which it is true both (1) that, if someone's life went in one of two ways, this would produce a greater sum total of local desire-fulfilment, but (2) that the other alternative is what he would globally prefer, whichever way his actual life went. (Summative Theories)*

1. **Objective List Theory:** certain things are good or bad for people, whether or not these people would want to have the good things, or to avoid the bad things. The good things might include moral goodness, rational activity, the development of one's abilities, having children and being a good parent, knowledge, and the awareness of true beauty. The bad things might include being betrayed, manipulated, slandered, deceived, being deprived of liberty or dignity, and enjoying either sadistic pleasure, or aesthetic pleasure in what is in fact ugly.

- coincides with the Global version of the Success Theory.
- On the Success Theory it is, for instance, bad for a person to be deceived if and because this is not what this person wants. The Objective List Theorist makes the reverse claim. People want not to be deceived because this is bad for them.
- distinction: P-H and ST give an account of self-interest, but OLT appeals directly to what it claims to be facts about value.

CONCLUSION:

Not decide which one is the most acceptable one.

> *Hedonists' view: assume the value of a whole is just the sum of the value of its parts. If we remove the part to which the Hedonist appeals, what is left seems to have no value, hence Hedonism is the truth.*

What is of value, or is good for someone, is to have both: to be engaged in these activities, and to be strongly wanting to be so engaged.

ESSAY 21: Nicomachean Ethics, Aristotle

尼各马可伦理学

BOOK I

Book I attempts to both define the subject matter itself and justify the method which has been chosen (in chapters 3, 4, 6 and 7). As part of this, Aristotle considers common opinions along with the opinions of poets and philosophers.

Who should study ethics, and how

Concerning accuracy and whether ethics can be treated in an objective way, Aristotle points out that the "things that are beautiful and just, about which politics investigates, involve great disagreement and inconsistency, so that they are thought to belong only to convention and not to nature". For this reason Aristotle claims it is important not to demand too much precision, like the demonstrations we would demand from a mathematician, but rather to treat the beautiful and the just as "things that are so for the most part". We can do this because people are good judges of what they are acquainted with, but this in turn implies that the young (in age or in character), being inexperienced, are not suitable for study of this type of political subject.

Defining happiness and the aim of the Ethics

The main stream of discussion starts in Chapter 1, from an assertion that all making, investigating (every *methodos*, like the *Ethics* itself), all deliberate actions and choice, all aim at some good. Aristotle points to be the fact that many aims are really only intermediate aims, and are desired only because they make the achievement of higher aims possible.

In chapter 2, Aristotle asserts that there is one highest aim, happiness, and it must be the same as politics should have, because what is best for an individual is less beautiful (kalos) and divine (theios) than what is good for a people (ethnos) or city (polis). The aim of political capacity should include the aim of all other pursuits, so that "this end would be the human good" a term which contrasts with Plato's references to "the Good itself". He concludes what is now known as Chapter 2 of Book 1 by stating that ethics ("our investigation" or methodos) is "in a certain way political".

Chapter 3 goes on to elaborate on exactness in its relation to the sought conclusions of actions. It is determined that the degree of exactness required in concluding arguments made for actions is not universally the same, and he goes on to explain that all actions, as with the conclusions made of them, vary in exactness and indeed for this reason are not universally true and therefore not universally applicable. It is for this reason that the continuing conclusions made throughout the processes of these examinations should be considered only in "outline" and that the premises on which the conclusions are drawn are on actions that hold the conclusions only usually.

Chapter 4 states that while most would agree to call the highest aim of humanity happiness (eudaimonia), and also to equate this with both living well and doing things well, there is dispute between people, and between the majority (hoi polloi) and "the wise". Chapter 5 distinguishes three distinct ways of life which different people associate with happiness.

- The slavish way of pleasure, which is the way the majority of people think of happiness.
- The refined and active way of politics, which aims at honor, honor itself implying the incompleteness of this way also, and the higher divinity of those who are wise and know and judge, and potentially honor, political people.
- The way of contemplation.

Aristotle also mentions two other possibilities that he argues can be put aside:

- Having virtue but being inactive, even suffering evils and misfortunes, which Aristotle says no one would consider unless they were defending a hypothesis. (As Sachs points out, this is indeed what Plato depicts Socrates doing in his Gorgias.)
- Money making, which Aristotle asserts to be a life based on aiming at what is pursued by necessity in order to achieve higher goals, an intermediate good.

Each of these three commonly proposed happy ways of life represents a target that people aim at for its own sake, just like they aim at happiness itself for its own sake.

Concerning honor, pleasure, and intelligence (nous) and also every virtue, though they lead to happiness, even if they did not we would still pursue them.

Happiness in life then, includes the virtues, and Aristotle adds that it would include self-sufficiency (autarkeia), not the self-sufficiency of a

hermit, but of someone with a family, friends and community. By itself this would make life choiceworthy and lacking nothing. In order to describe more clearly what happiness is like, Aristotle next asks what the work (ergon) of a human is. All living things have nutrition and growth as a work, all animals (according to the definition of animal Aristotle used) would have perceiving as part of their work, but what is more particularly human? The answer according to Aristotle is that it must involve articulate speech (logos), including both being open to persuasion by reasoning, and thinking things through. Not only will happiness involve reason, but it will also be an active being at work (energeia), not just potential happiness, and it will be over a lifetime, because "one swallow does not make a spring". The definition given is therefore:

> *the Good of man is the active exercise of his soul's faculties in conformity with excellence or virtue, or if there be several human excellences or virtues, in conformity with the best and most perfect among them. Moreover, to be happy takes a complete lifetime; for one swallow does not make spring (*from Rackham's translation)*

And because happiness is being described as a work or function of humans, we can say that just as we contrast harpists with serious harpists, the person who lives well and beautifully in this actively rational and virtuous way will be a "serious" (spoudaios) human.

As an example of popular opinions about happiness, Aristotle cites an "ancient one and agreed to by the philosophers". According to this opinion, which he says is right, the good things associated with the soul are most governing and especially good, when compared to the good things of the body, or good external things. Aristotle says that virtue, practical judgment and wisdom, and also pleasure, all associated with happiness, and indeed an association with external abundance, are all consistent with this definition.

If happiness is virtue, or a certain virtue, then it must not just be a condition of being virtuous, potentially, but an actual way of virtuously "being at work" as a human. For as in the Ancient Olympic Games, "it is not the most beautiful or the strongest who are crowned, but those who compete". And such virtue will be good, beautiful and pleasant, indeed Aristotle asserts that in most people different pleasures are in conflict with each other while "the things that are pleasant to those who are passionately devoted to what is beautiful are the things that are pleasant by nature and of this sort are actions in accordance with virtue". External goods are also necessary in such a virtuous life, because a person who lacks things such as good family and friends might find it difficult to be happy.

From defining happiness to discussion of virtue: introduction to the rest of the Ethics

Aristotle asserts that we can usefully accept some things which are said about the soul (clearly a cross reference to Plato again), including the division of the soul into rational and irrational parts, and the further division of the irrational parts into two parts also:

— One irrational part of the human soul is "not human" but "vegetative" and at most work during sleep, when virtue is least obvious.

— A second irrational part of the human soul is however able to share in reason in some way. We see this because we know there is something "desiring and generally appetitive" in the soul which can on different occasions in different people either oppose reason, or obey it—thus being rational just as we would be rational when we listen to a father being rational.

The virtues then will be similarly divided, into intellectual (dianoetic) virtues, and the virtues of character (ethical or moral virtues) pertaining to the irrational part of the soul which can take part in reason.[27]

These virtues of character, or "moral virtues" as they are often translated, become the central topic in Book II. The intellectual aspect of virtue will be discussed in Book VI.

Book II: That virtues of character can be described as means

Aristotle says that whereas virtue of thinking needs teaching, experience and time, virtue of character (moral virtue) comes about as a consequence of following the right habits. According to Aristotle the potential for this virtue is by nature in humans, but whether virtues come to be present or not is not determined by human nature.

Trying to follow the method of starting with approximate things gentlemen can agree upon, and looking at all circumstances, Aristotle says that we can describe virtues as things which are destroyed by deficiency or excess. Someone who runs away becomes a coward, while someone who fears nothing is rash. In this way the virtue "bravery" can be seen as depending upon a "mean" between two extremes. (For this reason, Aristotle is sometimes considered a proponent of a doctrine of a "golden mean".) People become habituated well by first performing actions which are virtuous, possibly because of the guidance of teachers or experience, and in turn these habitual actions then become real virtue where we choose good actions deliberately.

According to Aristotle character properly understood, meaning one's virtue or vice, is not just any tendency or habit but something which affects when we feel pleasure or pain. A virtuous person feels pleasure at the most beautiful or noble (kalos) actions. A person who is not virtuous will often find his or her perceptions of what is most pleasant to be misleading. For this reason, any concern with virtue or politics requires consideration of pleasure and pain. When a person does virtuous actions, for example by chance, or under advice, they are not yet necessarily a virtuous person. It is not like in the productive arts, where the thing being made is what is judged as well made or not. To truly be a virtuous person, one's virtuous actions must meet three conditions: (a) they are done knowingly, (b) they are chosen for their own sakes, and

(c) they are chosen according to a stable disposition (not at a whim, or in any way that the acting person might easily change his choice about). And just knowing what would be virtuous is not enough. According to Aristotle's analysis, there are three kinds of things which come to be present in the soul that virtue is: a feeling (pathos), an inborn predisposition or capacity (dunamis), or a stable disposition which has been acquired (hexis). In fact, it has already been mentioned that virtue is made up of hexeis, but on this occasion the contrast with feelings and capacities is made clearer—neither are chosen, and neither are praiseworthy in the way that virtue is.

Comparing virtue to productive arts (technai) as with arts, virtue of character must not only be the making of a good human, but also the way in which a human does his own work well. And being skilled in an art can also be described as a mean between excess and deficiency: when they are well done we say that we would not want to take away or add anything from them. But Aristotle points to a simplification in this idea of hitting a "mean". In terms of what is best, we aim at an extreme, not a mean, and in terms of what is base, the opposite.

Chapter *7* turns from general comments to specifics. A list of virtues and vices of character are given which will be discussed in Books II and III. As Sachs points out (2002, p. 30) it appears that the list is not especially fixed, because it differs between the Nicomachean and Eudemian Ethics, and also because Aristotle repeats several times that this is a rough outline.

Aristotle also mentions some "mean conditions" involving feelings: a sense of shame is sometimes praised, or said to be in excess or deficiency. Righteous indignation (Greek: nemesis) is a sort of mean between joy at the misfortunes of others and envy. Aristotle says that such cases will need to be discussed later, before the discussion of Justice in Book V, which will also require special discussion. But the Nicomachean Ethics only discusses the sense of shame at that point, and not righteous indignation (which is however discussed in the Eudemian Ethics Book VIII).

In practice Aristotle explains that people tend more by nature towards pleasures, and therefore see virtues as being relatively closer to the less obviously pleasant extremes. While every case can be different, given the difficulty of getting the mean perfectly right it is indeed often most important to guard against going the pleasant and easy way. However this rule of thumb is shown in later parts of the Ethics to apply mainly to some bodily pleasures, and is shown to be wrong as an accurate general rule in Book X.

BOOK I

Summary

Every human activity aims at some end that we consider good. The highest ends are ends in themselves, while subordinate ends may only be means to higher ends. Those highest ends, which we pursue for their own sake, must be the supreme Good.

The study of the Good is part of political science, because politics concerns itself with securing the highest ends for human life. Politics is not a precise science, since what is best for one person may not be best for another. Consequently, we can aim at only a rough outline of the Good.

Everyone agrees that the supreme Good is happiness, but people disagree over what constitutes happiness. Common people equate happiness with sensual pleasure: this may be sufficient for animals, but human life has higher ends. Others say that receiving honors is the greatest good, but honors are conferred as recognition of goodness, so there must be a greater good that these honors reward. Plato's Theory of Forms suggests that there is a single Form of Good and that all good things are good in the same way. This theory seems flawed when we consider the diversity of things we call "good" and the diversity of ways in which we consider goodness. Even if there were a single unifying Form of Good, our interest is in the practical question of how to be good, so we should concern ourselves not with this abstract concept but with the practical ends we can actually pursue in everyday life.

Happiness is the highest good because we choose happiness as an end sufficient in itself. Even intelligence and virtue are not good only in themselves, but good also because they make us happy.

We call people "good" if they perform their function well. For instance, a person who plays the flute well is a good flutist. Playing the flute is the flutist's function because that is his or her distinctive activity. The distinctive activity of humans generally—what distinguishes us from plants and animals—is our rationality. Therefore, the supreme Good should be an activity of the rational soul in accordance with virtue. This definition aligns with popular views of happiness, which see the happy person as virtuous, rational, and active.

When talking about happiness, we consider a person's life as a whole, not just brief moments of it. This raises the paradoxical suggestion that a person can be considered happy only after death, that is, once we can examine the person's life as a whole. However, a good person will always behave in a virtuous manner. Even faced with great misfortune, a good person will bear himself or herself well and will not descend into mean-spiritedness. Once a person has died, according to Aristotle, posthumous honors or dishonors and the behavior of his descendants might affect his happiness somewhat, but to no great extent.

We can divide the soul into an irrational and a rational part. The irrational soul has two aspects: the vegetative aspect, which deals with

nutrition and growth and has little connection to virtue; and the appetitive aspect, which governs our impulses. The rational part of the soul controls these impulses, so a virtuous person with greater rationality is better able to control his or her impulses.

Analysis

Much confusion about Aristotle's work comes not from Aristotle's lack of clarity, but from an imprecision in translation. Ancient Greek is quite different from the English language, and more important, the ancient Greeks lived in a very different culture that used concepts for which there are no exact English translations.

One central concept of the *Ethics* is *eudaimonia,* which is generally translated as "happiness." While happiness is probably the best English word to translate *eudaimonia,* the term also carries connotations of success, fulfillment, and flourishing. A person who is *eudaimon* is not simply enjoying life, but is enjoying life by living successfully. One's success and reputation, unlike one's emotional well-being, can be affected after death, which makes Aristotle's discussion of *eudaimonia* after death considerably more relevant.

That happiness should be closely connected to success and fulfillment reflects an important aspect of social life in ancient Greece. The identity of Greek citizens was so closely linked to the city-state to which they belonged that exile was often thought of as a fate worse than death. There was no distinction between the public and private spheres as exists in the modern world. Consequently, happiness was not thought of as a private affair, dependent on individual emotional states, but as a reflection of a person's position within a city-state. A person who inhabits a proper place in the social structure and who appropriately fulfills the duties and expectations of that place is "happy" because, for the Greeks, happiness is a matter of living—not just feeling—the right way.

Aristotle treats happiness as an activity, not as a state. He uses the word *energeia,* which is the root of our word *energy*, to characterize happiness. The point is that happiness consists of a certain way of life, not of certain dispositions. In saying that happiness is an *energeia,* he contrasts happiness with virtue, which he considers a *hexis,* or state of being. Possessing all the right virtues disposes a person to live well, while happiness is the activity of living well, which the virtuous person is inclined toward.

[T]he good for man is an activity of the soul in accordance with virtue, or if there are more kinds of virtue than one, in accordance with the best and most perfect kind.

(See Important Quotations Explained)

The very idea of living well might seem a bit odd as Aristotle formulates it. In particular, he talks about living well as performing the function of "being human" well, analogous to the good flutist performing the function of playing the flute well. It may seem that Aristotle has confused the practical and the moral: being a good flutist is a practical matter of study and talent, while no such analogy holds for morality. Being a good person surely is not a skill one develops in the same manner as flute playing. But this objection rests on a misunderstanding due to a

difficulty in translation. The Greek word *ethos* translates as "character," and the concerns of the *Ethics* are not with determining what is right and wrong, but with how to live a virtuous and happy life.

We should also note the importance of the concept of *telos,* which we might translate as "end" or "goal." The first sentence of the *Ethics* tells us that every activity aims at a certain *telos.* For instance, one might go to the gym with the *telos* of becoming fitter. When Aristotle identifies happiness as the highest goal, he is claiming that happiness is the ultimate *telos* of any action. We might understand this idea of an ultimate *telos* by imagining the child who constantly asks, "why?":

"Why are you going to the gym?"

"To become fitter."

"Why do you want to become fitter?"

"So that I'll be healthier."

"Why do you want to be healthy?"

"So that I'll live longer and have more energy."

"Why do you want a long and energetic life?"

"Because that makes for a happy life."

"Why do you want a happy life?"

"I just do."

Every activity has a *telos,* which is an answer to the question, *Why are you doing this?* Happiness is the ultimate *telos* because there is no further *telos* beyond happiness and because the ultimate goal of all our other activities is happiness.

For Aristotle, the soul, or *psuche* (the root of our word *psychology*), is simply that which distinguishes living things from nonliving things. All living things have a nutritive soul, which governs bodily health and growth. Animals and humans differ from plants in having an appetitive soul, which governs movement and impulse. Humans differ from animals in also having a rational soul, which governs thought and reason. Because rationality is the unique achievement of humans, Aristotle sees rationality as our *telos*: in his view, everything exists for a purpose, and the purpose of human life is to develop and exercise our rational soul. Consequently, a human can "be human" well by developing reason in the way that a flutist can be a good flutist by developing skill with the flute.

BOOK II

Summary

There are two kinds of virtue: intellectual and moral. We learn intellectual virtues by instruction, and we learn moral virtues by habit and constant practice. We are all born with the potential to be morally virtuous, but it is only by behaving in the right way that we train ourselves to be virtuous. As a musician learns to play an instrument, we learn virtue by practicing, not by thinking about it.

Because practical circumstances vary a great deal, there are no absolute rules of conduct to follow. Instead, we can only observe that right conduct consists of some sort of mean between the extremes of deficiency and excess. For instance, courage consists in finding a mean between the extremes of cowardice and rashness, though the appropriate amount of courage varies from one situation to another.

An appropriate attitude toward pleasure and pain is one of the most important habits to develop for moral virtue. While a glutton might feel inappropriate pleasure when presented with food and inappropriate pain when deprived of food, a temperate person will gain pleasure from abstaining from such indulgence.

Aristotle proposes three criteria to distinguish virtuous people from people who behave in the right way by accident: first, virtuous people know they are behaving in the right way; second, they choose to behave in the right way for the sake of being virtuous; and third, their behavior manifests itself as part of a fixed, virtuous disposition.

Virtue is a disposition, not a feeling or a faculty. Feelings are not the subject of praise or blame, as virtues and vices are, and while feelings move us to act in a certain way, virtues dispose us to act in a certain way. Our faculties determine our capacity for feelings, and virtue is no more a capacity for feeling than it is a feeling itself. Rather, it is a disposition to behave in the right way.

We can now define human virtue as a disposition to behave in the right manner and as a mean between extremes of deficiency and excess, which are vices. Of course, with some actions, such as murder or adultery, there is no virtuous mean, since these actions are always wrong. Aristotle lists some of the principle virtues along with their corresponding vices of excess and deficiency in a table of virtues and vices. Some extremes seem closer to the mean than others: for instance, rashness seems closer to courage than to cowardice. This is partly because courage is more like rashness than cowardice and partly because most of us are more inclined to be cowardly than rash, so we are more aware of being deficient in courage.

Aristotle suggests three practical rules of conduct: first, avoid the extreme that is farther from the mean; second, notice what errors we are

particularly susceptible to and avoid them diligently; and third, be wary of pleasure, as it often impedes our judgment.

Analysis

"Virtue" is the most common translation of the Greek word *arete,* though it is occasionally translated as "excellence." Virtue is usually an adequate translation in the *Ethics* because it deals specifically with human excellence, but *arete* could be used to describe any kind of excellence, such as the sharpness of a knife or the fitness of an athlete. Just as a knife's excellence rests in its sharpness, a person's excellence rests in living according to the various moral and intellectual virtues.

Aristotle describes virtue as a disposition, distinguishing it not only from feelings and faculties, but also (less explicitly) from activities. Aristotle calls happiness an activity, or *energeia,* in Book I, meaning that happiness is not an emotional state but a way of life. Happiness is exhibited not in how we are but in how we act. Virtue, by contrast, is a disposition, or *hexis,* meaning that it is a state of being and not an activity. More precisely, virtue is the disposition to act in such a way as to lead a happy life.

Without virtue, we cannot be happy, though possessing virtue does not in itself guarantee happiness. In Book I, Chapter 8, Aristotle points out that those who win honors at the Olympic Games are not necessarily the strongest people present but rather the strongest people who actually compete. Perhaps one of the spectators is strong-er than all of the competitors, but this spectator has no right to win honors. Similarly, a person might have a virtuous disposition but will not lead a happy life unless he or she acts according to this disposition.

It may seem odd to us that Aristotle at no point argues for what dispositions should be considered virtuous and which vicious. The need for justification seems even more pressing in the modern world, where our views on virtue and vice may not entirely agree with Aristotle's.

However, it is not Aristotle's intention to convince us of what is virtuous, and he differs from most modern moral philosophers in placing very little emphasis on rational argument in moral development. Instead, as he argues at the beginning of Book II, learning virtue is a matter of habit and proper training. We do not become courageous by learning why courage is preferable to cowardice or rashness, but rather by being trained to be courageous. Only when we have learned to be instinctively courageous can we rightly arrive at any reasoned approval of courage. Recalling that *arete* may refer to any form of excellence, we might draw an analogy between learning courage and learning rock-climbing. We learn to become good rock-climbers through constant practice, not through reasoned arguments, and only when we have become good rock-climbers and appreciate firsthand the joys of rock-climbing can we properly understand why rock-climbing is a worthwhile activity.

Aristotle's conception of virtue as something learned through habit rather than through reasoning makes a great deal of practical sense. We can generally trace unpleasantness to the circumstances in which a person grew up, and it is difficult to make an unpleasant person pleasant simply by providing reasons for behaving more pleasantly.

The virtues Aristotle lists, then, reflect the commonly held values of a properly raised, aristocratic Athenian. If we disagree with Aristotle's

choices of virtues, we are unlikely to find a compelling argument in his work to change our mind: by Aristotle's own admission, reasoning is unlikely to teach us to appreciate virtue if we have not been raised with the right habits.

One of the most celebrated and discussed aspects of Aristotle's *Ethics* is his Doctrine of the Mean, which holds that every virtue is a mean between the vicious extremes of excess and deficiency. This is not a strict rule, as Aristotle himself points out: there is no precise formula by which we can determine exactly where this mean lies, largely because the mean will vary for different people.

That there should be no fixed rule to determine where the mean lies is a direct consequence of his doctrine that virtue is something learned through habit, not through reason. If we could reason our way into virtue, we might be able to set out precise rules for how to behave in different situations. According to Aristotle's view, however, a virtuous person is naturally inclined to choose the correct behavior in any situation without appealing to rules or maxims.

In Book I, Chapter 3, and Book II, Chapter 2, Aristotle warns us that our inquiry is at best an imprecise one. Bearing in mind that virtue for Aristotle is a set of innate dispositions, not a reasoned set of rules, we can understand these warnings to be more than simple hedges. Aristotle is not avoiding precision but saying that precision is impossible because there are no fixed rules of conduct that we can follow with confidence.

ESSAY 22: Slections from Grounding for the Metaphysics of Morals, Immanuel Kant

道德形而上学的基础

Section 1: Transition from the ordinary rational knowledge of Morality to the Philosophical:

In this section, Kant presents and defends his idea that what he calls the "categorical imperative" serves as the basis for morality.

定言令式

A good will is good not because of what it effects or accomplishes, nor because of its fitness to attain some proposed end; it is good only through its willing, i.e., it is good in itself.

• The purpose of Section 1 is to "proceed analytically from ordinary knowledge to a determination of the supreme principle" (392). Which is to say that Kant intends here to move from our ordinary ways of thinking about morality, analyzing them to discover the principles which lie behind them.

• Here, Kant is trying to prove something antecedent to the fact that we have moral obligations - he is trying to show what it is that he has to establish in order to show that morality is possible.

• Kant's starting point in his argument (he assumes that everyone would agree with this belief) is that a "good will" is the only thing to which we attribute unconditional moral value. What he means is that the 'good will' is the only thing which has value completely independently of anything external to it, and which it therefore has in all circumstances, independent of contingent empirical facts.

• Kant thinks that we cannot detract from the value of an action done from a good will, even if that actions turns out to be unsuccessful. The value of such an action is independent of "what it effects or accomplishes" (394). (Compare this view with the consequential structure of utilitarianism).

• Thus, Kant's project becomes that of "elucidating" the concept of a good will (397): Kant is going to find out what principle the person of

good will acts on, in order to establish what the moral law tells us to do.

• Kant focuses on actions done from duty. Duty is the good will operating "though with certain subjective restrictions and hindrances, which ... far from hiding a good will and making it unrecognizable, rather bring it out by contrast and make it shin forth more brightly." (397). By these 'subjective hindrances', Kant has in mind the person who has other motives which would mean that, in the absence of a sense of duty, that person would not be motivated at all to perform the morally right action.

• Kant identifies three kinds of motivation for action:

1. Duty - you perform the action because you think that it's the right thing to do
2. Immediate Inclination - you simply enjoy doing actions of a particular sort
3. Instrumental Inclination - you perform the action because of some independent end which it serves

• Kant thinks that right actions performed from duty have a special value which right actions performed from one of the other kinds of motivation lack.

• Kant gives the example of the 'prudent merchant', who acts fairly towards his customers because this will secure his reputation, but not for its own sake. (The merchant has type-3 motivation).

• More controversially, Kant thinks that the actions of the naturally beneficent or sympathetic person, who does not act from duty, but does good because he is naturally inclined towards doing the right thing, and enjoys doing so, also has no moral worth, insofar as it is not action from duty, and hence does not evince a good will. (i.e. type-2 motivation) "I maintain that in such a case an action of this kind, however dutiful and amiable it may be, has nevertheless no true moral worth." (398).

• Kant thinks that, for an action to be morally worthy, it has to be performed for the reason that you think that it is required of you (Duty) - if it is simply the case that it pleases you to do the morally right thing, then that is not enough for moral worth, as your action is not independent of the contingent fact that you happen to have certain preferences which incline you towards performing it.

• So, Kant now has an account of what makes morally worthy actions have their special worth. They get their moral worth from the fact that the person who performs them acts from respect for moral law, and not through any reason independent of that moral law.

Summary

The one thing in the world that is unambiguously good is the "good will." Qualities of character (wit, intelligence, courage, etc.) or qualities of good fortune (wealth, status, good health) may be used to either good or bad purposes. By contrast, a good will is intrinsically good--even if its efforts fail to bring about positive results.

It is a principle of the composition of natural organisms that each of their purposes is served by the organ or faculty most appropriate to that purpose. The highest purposes of each individual are presumably self-preservation and the attainment of happiness. Reason does not appear to be as well suited as instinct for these purposes. Indeed, people with a refined capacity for reason are often less happy than the masses. As a result, refined people often envy the masses, while common people view reason with contempt. The fact is that reason serves purposes that are higher than individual survival and private happiness. Reason's function is to bring about a will that is good *in itself,* as opposed to good for some particular purpose, such as the attainment of happiness.

The specific obligations of a good will are called "duties." We may make three general propositions about duty. First, actions are genuinely good when they are undertaken for the sake of duty alone. People may act in conformity with duty out of some interest or compulsion other than duty. For instance, a grocer has a duty to offer a fair price to all customers, yet grocers abide by this duty not solely out of a sense of duty, but rather because the competition of other grocers compels them to offer the lowest possible price. Similarly, all people have a duty to help others in distress, yet many people may help others not out of a sense of duty, but rather because it gives them pleasure to spread happiness to other people. A more genuine example of duty would be a person who feels no philanthropic inclination, but who nonetheless works to help others because he or she recognizes that it is a duty to do so.

The second proposition is that actions are judged not according to the purpose they were meant to bring about, but rather by the "maxim" or principle that served as their motivation. This principle is similar to the first. When someone undertakes an action with no other motivation than a sense of duty, they are doing so because they have recognized a moral principle that is valid *a priori.* By contrast, if they undertake an action in order to bring about a particular result, then they have a motivation beyond mere duty.

The third proposition, also related to the first two, is that duties should be undertaken out of "reverence" for "the law." Any organism can act out of instinct. Chance events could bring about positive results. But only a rational being can recognize a general moral law and act out of respect for it. The "reverence" for law that such a being exhibits (this is explained in Kant's footnote) is not an emotional feeling of respect for the greatness of the law. Rather, it is the moral motivation of a person who recognizes that the law is an imperative of reason that transcends all other concerns and interests.

Since particular circumstances and motivations cannot be brought into the consideration of moral principles, the moral "law" cannot be a specific stipulation to do or not do this or that particular action. Rather, the moral law must be applicable in all situations. Thus the law of morality is that we should act in such a way that we could want the maxim (the motivating principle) of our action to become a universal law.

Giving a false promise is an example of an action that violates this moral law. Some people might reason that they should be permitted to lie in order to escape a difficult situation. Conversely, some people might reason that they should not lie because in doing so they might create still greater difficulties for themselves in the future. In both cases, the motivating consideration is a fear of consequences, not pure respect for duty. Applying the moral law reveals that lying can never be a universal law. If everyone were to give false promises, then there would be no such thing as a promise.

Although most people are not aware of the moral law in any conscious sense, even untrained minds show a remarkable ability to abide by it in practice. People's intuitive sense for theoretical matters is generally poor. By contrast, their intuitions in the field of practical reason--in other words, their intuitions about morality--are generally correct. For instance, people generally recognize that moral concerns should not include physical ("sensuous") motivations. Nevertheless, a philosophical understanding of morals is important, because untrained minds may be deceived and distracted by non-moral needs, concerns, and desires.

Commentary

Since Kant's argument in this chapter is complex, it may be helpful to paraphrase it in a more compressed form. Kant starts from the presumption that an action is moral if and only if it is intrinsically good--good "in itself," as he puts it. This view has two main implications. First, moral actions cannot have impure motivations. Otherwise, the action would be based on some secondary motivation, and not on the intrinsic goodness of the action. Second, moral actions cannot be based on consideration of possible outcomes. Otherwise, the action would not be good in itself, but would instead be good in that it brought about a particular outcome.

If we can consider neither motivating circumstances nor intended outcomes, then we need to find a principle with universal validity--a principle that is valid no matter what issue we are considering. The only principles that fit this criterion are the *a priori* principles of reason--that is, the principles of logic that we have to follow if our statements are to make sense.

One fundamental principle of logic is the principle of non-contradiction: statements don't make sense if they contradict themselves. Kant's moral law is based on this principle of non-contradiction. In order for your action to be moral, he argues, it must be good in itself. In order for it be good in itself, it must make sense in pure logical terms. In order for it to make sense, it must not contradict itself. If you lie but expect other people to believe you, you contradict yourself. Your motivation lacks universal validity and is therefore immoral.

At the end of the chapter, Kant argues that his analysis of the moral law amounts, in effect, to a formalization of a moral sense that we already use intuitively. He argues that a more conscious understanding of the principles of our moral sense can help us to behave more morally.

Given the complexity of his argument, it may seem surprising that he believes he is only teaching us what we already know. His claim may seem less surprising if we recognize that his moral law is fundamentally the same as the Biblical teaching that we should "do unto others as we would have done unto us." Kant argues that we violate rational principles of morality when we contradict ourselves, and that we contradict ourselves when we act in a way that we would not want others to imitate. In practice, his doctrine amounts to a doctrine of respect for others.

The major criticism of Kant's approach is that it is too abstract to be useful. The nineteenth-century philosopher Hegel is generally credited with developing this argument against Kant. Hegel argued that our thinking is structured by the beliefs, institutions, and traditions of the society in which we live. In criticism of Kant, he pointed out that you cannot know what actions will appear self-contradictory to people unless you know something about their society.

Take the prohibition against theft, for example. We live in a world of property. In our world, it is contradictory to steal, because when you steal you expect others to recognize your ownership of what you have stolen even though you failed to respect the ownership of the person who originally possessed it. So far, Kant's analysis holds. Yet we can imagine a world without property rights, a world where everything is collectively owned. In such a world, there would be no such thing as theft because there would be no such thing as personal property.

The same analysis can be applied to nearly every moral principle. In our society, it is unethical to cheat on your spouse, because you contradict yourself when accept the marriage vows of your spouse and yet break those vows yourself. Yet we could imagine a world with different family institutions where affairs might not be considered unethical. Similarly (to use Kant's example), it is unethical in our society to make false promises. In our society, there is such a thing as a promise, and when people make promises we expect that they will keep them. But lying might mean something different in a society with different expectations.

According to Hegel's analysis, Kant is correct to recognize that the principle of non-contradiction is an element in moral thinking, but he is wrong to think that we can develop moral principles without considering the circumstances of our world. Morality is not something for automatons living a life of pure rational thought. It is a consideration for human beings who must sometimes subordinate their personal interests to the basic principles of their community.

In defense of Kant against Hegel, some philosophers (##Kierkegaard##, for instance) have criticized Hegel for overemphasizing the role that social institutions play in forming our beliefs. By some accounts, Kant has the advantageous of allowing us greater freedom in reasoning about which morals make sense to us, independent of the society around us. We will continue to consider this and other views on Kant as we consider his further arguments in Chapters 2 and 3.

As a small side note, it may be of interest to note that Kant wrote the *Grounding for the Metaphysics of Morals* over half a century before ##Charles Darwin## formulated his theory of evolution by natural selection. From a modern-day perspective, Kant's statement that an organism's needs are generally served by the most-suited organ might seem a little strange. An evolutionary biologist would say that our organs and

faculties have developed over time in order to serve the needs of survival. According to this perspective, we wouldn't have organs or faculties unless they served our survival needs (or had served those needs at some time); the point is that our organs and faculties should work, not that they perform the tasks for which they are best suited. Kant's outdated view of nature is not of critical importance to his argument, however, so this is not a major problem. It may be interesting nonetheless to observe that ideas about instinct and self- preservation were established long before Darwin included them in his theory.

ESSAY 23: Selections from Utilitarianism, John Stuart Mill

Chapter 2: What Utilitarianism Is

Summary

Mill attempts to reply to misconceptions about utilitarianism, and thereby delineate the theory. Mill observes that many people misunderstand utilitarianism by interpreting utility as in opposition to pleasure. In reality, utility is defined as pleasure itself, and the absence of pain. Thus another name for utility is the Greatest Happiness Principle. This principle holds that "actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure." Pleasure and the absence of pain are, by this account, the only things desirable as ends in themselves, the only things inherently "good." Thus, any other circumstance, event, or experience is desirable only insofar as it is a source for such pleasure; actions are good when they lead to a higher level of general happiness, and bad when they decrease that level.

The next criticism Mill takes on is the claim that it is base and demeaning to reduce the meaning of life to pleasure. To this Mill replies that human pleasures are much superior animalistic ones: once people are made aware of their higher faculties, they will never be happy to leave them uncultivated; thus happiness is a sign that we are exercising our higher faculties. It is true that some pleasures may be "base"; however, this does not mean that all of them are: rather, some are intrinsically more valuable than others. When making a moral judgment on an action, utilitarianism thus takes into account not just the quantity, but also the quality of the pleasures resulting from it.

Mill delineates how to differentiate between higher- and lower-quality pleasures: A pleasure is of higher quality if people would choose it over a different pleasure even if it is accompanied by discomfort, and if they would not trade it for a greater amount of the other pleasure.

Moreover, Mill contends, it is an "unquestionable fact" that, given equal access to all kinds of pleasures, people will prefer those that appeal to their "higher" faculties. A person will not choose to become an animal, an educated person will not choose to become ignorant, and so on. Even though a person who uses higher faculties often suffers more in life (hence the common dictum "ignorance is bliss"), he would never choose a lower existence, preferring instead to maintain his dignity.

Another misconception about utilitarianism stems from a confusion of happiness with contentment. People who employ higher faculties are often less content, because they have a deeper sense of the limitations of the world. However, their pleasure is of a higher character than that of an animal or a base human. Mill writes, "It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied. And if the fool, or the pig, are of a different opinions, it is because they only know their side of the question." Thus the people best qualified to judge a pleasure's quality are people who have experienced both the higher and the lower.

Furthermore, Mill observes that even if the possession of a "noble character" brought less happiness to the individual, society would still benefit. Thus, because the greatest happiness principle considers the total amount of happiness, a noble character, even if it is less desirable for the individual, is still desirable by a utilitarian standard.

Commentary

This chapter provides the definition of utilitarianism. There are a few important aspects of this definition. First, it presents utility, or the existence of pleasure and the absence of pain, as both the basis of everything that people desire, and as the foundation of morality. However, utilitarianism does not say that it is moral for people simply to pursue what makes them personally happy. Rather, morality is dictated by the greatest happiness principle; moral action is that which increases the total amount of utility in the world. Pursuing one's own happiness at the expense of social happiness would not be moral under this framework.

The most significant aspect of this section, however, is Mill's discussion of the higher and lower pleasures. Over the years, utilitarianism's critics have often objected that it tries to compare things that are fundamentally incommensurable, by artificially computing the amount of utility they bring. For example, by reducing the value of an experience or action to the utility, or pleasure, inherent in them, utilitarianism "cheapens" certain experiences: is it fair to compare eating ice cream to reading *War and Peace,* based on the pleasure each brings? In this chapter, Mill tries to address this concern. He argues that utility is not simply a measurement of the psychological feeling of pleasure; rather, there are different qualities of pleasure, and only people with a broad range of experiences can dictate which pleasures are of a higher quality. Thus all actions and experiences are not judged by one reductive standard, but rather according to a variety of different qualities of pleasure in correspondence with the type of experience. Higher pleasures would be weighted heavily by utilitarianism, and Mill argues that they are therefore not cheapened by the utility measurement.

It is important, then, to consider whether Mill has adequately responded to criticisms about incommensurable pleasures; is Mill's explana-

tion complete? We still might ask what it is that makes some pleasures "superior" to others. When we say that a pleasure is "higher," what do we really mean? That it is more educational? Appreciated only by those with good taste? Appreciated only by the intelligent? Utility is supposed to be a foundational measurement, but perhaps to acknowledge the existence of higher and lower pleasures is to admit a standard of measurement other than mere pleasure. How might Mill respond to this objection?

Summary

Having responded to the objection that utilitarianism glorifies base pleasures, Mill spends the rest of this chapter presenting and responding to other criticisms of utilitarianism.

One such objection is that happiness couldn't be the rational aim of human life, because it is unattainable. Furthermore, people can exist without happiness, and all virtuous people have become virtuous by renouncing happiness.

First, Mill replies that it is an exaggeration to state that people cannot be happy. He contends that happiness, when defined as moments of rapture occurring in a life troubled by few pains, is indeed possible, and would be possible for almost everybody if educational and social arrangements were different. The major sources of unhappiness are selfishness and a lack of mental cultivation. Thus, it is fully within most people's capabilities to be happy, if their education nurtures the appropriate values. Furthermore, most of the evils of the world, including poverty and disease, can be alleviated by a wise and energetic society devoted to their elimination.

Next, Mill addresses the argument that the most virtuous people in history are those who have renounced happiness. He admits this is true, and he admits that there are martyrs who give up their happiness. However, Mill argues that martyrs must sacrifice happiness for some greater end--and what else could this be but the happiness of other people? The sacrifice is made so that others will not have to make similar sacrifices; implicit in the sacrifice is the value of others' happiness. Mill admits that the willingness to sacrifice one's happiness for that of others is the highest virtue. Furthermore, he says that to maintain an attitude of such willingness is actually the best chance of gaining happiness, because it will lead a person to be tranquil about his life and prospects. He specifies, however, that while utilitarians value sacrificing one's good for the good of others, they do not think that the sacrifice is in itself a good. It is a good insofar as it promotes happiness, but is not a good if it does not promote happiness.

Mill observes that the utilitarian's standard for judging an act is the happiness of *all* people, not of the agent alone. Thus, a person must not value his own happiness over the happiness of others; and law and education help to instill this generosity in individuals. However, this does not mean that people's motives must only be to serve the greatest good; indeed, utilitarianism is not concerned with the motives behind an action; the morality of an action depends on the goodness of its result only. Moreover, in most aspects of everyday life, a person will not be affecting large numbers of other people, and thus need not consider his or her actions in relation to the good of all, but only to the good of those

involved. It is only the people who work in the public sphere and affect many other people who must think about public utility on a regular basis.

Another criticism of utilitarianism is that it leaves people "cold and unsympathizing," as it is concerned solely with the consequences of people's actions, and not on the individuals as moral or immoral in themselves. First, Mill replies that if the criticism is that utilitarianism does not let the rightness or wrongness of an action be affected by the kind of person who performs the action, then this is a criticism of all morality: All ethical standards judge actions in themselves, without considering the morality of those who performed them. However, he says that if the criticism is meant to imply that many utilitarians look on utilitarianism as an exclusive standard of morality, and fail to appreciate other desirable "beauties of character," then this is a valid critique of many utilitarians. He says that it is a mistake to only cultivate moral feelings, to the exclusion of the sympathies or artistic understandings, a mistake moralists of all persuasions often make. However, he does say that if there is to be a mistake of priorities, it is preferable to err on the side of moral thinking.

Mill then presents a few more misunderstandings about utilitarian theory, which he declares are obviously wrong but which many people nonetheless believe. First, utilitarianism is often called a godless doctrine, because its moral foundation is the human happiness, and not the will of God. Mill replies that the criticism depends on what we see to be the moral character of God; for if God desires the happiness of all His creatures, then utilitarianism is more religious than any other doctrine. A utilitarian believes that God's revealed truths about morality will fit with utilitarian principles. Furthermore, many moralists, not simply utilitarians, have believed that we need an ethical doctrine, carefully followed, in order to understand the will of God in the first place.

Secondly, utilitarianism is often conflated with Expediency, and therefore considered immoral. However, "expedient" usually refers to acting against what is right for the sake of personal interest or short-term goals. Thus, instead of being useful, this meaning of expediency is actually harmful. Mill would argue that hurting society is not truly expedient, and that to act against society's interests is to be an enemy of morality.

Many critics hold that prior to taking action, there is often not enough time to weigh its effects on general utility. Mill dismisses this, saying that such a claim is akin to saying that we can't guide our conduct by Christianity because we can't read the Bible every time we had to act. He asserts that we have had the entire history of human existence within which to learn the tendencies of actions to lead to particular results. There is a great deal of consensus about what is useful, and we have the capacity to impart this knowledge to children too. This is not to say that received ethics are always correct, and there is still much to learn about the effects of actions on general happiness. However, people need not reapply the first principles to an action each time they perform it. All rational people go through life with their minds made up on certain basic questions of right and wrong.

Finally, utilitarianism is criticized as too allowing, as underestimating the immoral tendencies of human nature. For example, it is argued

that a utilitarian will make his own case an exception to the rules, and will be tempted to justify breaking the rules by simply saying that a given action increases utility. However, Mill says this problem is not limited to utilitarian theories. All creeds must have exceptions, because the need for exceptions is part of the reality of human life. Having a standard of utility to invoke is better than having no standard at all.

Commentary

ne of Mill's most common replies to objections about utilitarianism is that the given critique is not unique to utilitarianism, that any ethical theory would have such limitations. What are the strengths and weaknesses of this tactic? Does it really satisfy Mill's stated objective, to dispel misconceptions about his theory? Might such a reply undermine all ethical theories?

Mill makes some of his most controversial arguments in this section, and it is important to look closely at his arguments and assumptions. There is not an obvious right or wrong answer in this debate, but it may be helpful to think about some of the areas where Mill's argument is most commonly attacked. Mill observes that utilitarianism is concerned with increasing the amount of general happiness, not with increasing any one person's happiness. One common criticism of this concept is that by basing morality on the general good, utilitarianism fails to appreciate the importance of the individual. In dealing with this debate, it is useful to recognize a difference of perspective. Mill takes an impersonal perspective, where morality is impartial. One could, however, argue that morality should be subject-oriented, or interpeersonal. Another contentious point is Mill's argument that individuals' motives do not matter in morality. Is an action fundamentally different if it is performed for good or bad reasons? Mill would argue it is not. Finally, Mill argues that sacrificing happiness is only desirable if it will lead to more happiness generally. He rejects the value of sacrifice in itself. However, many people do see value in an ascetic life, independent of the consequences it produces. This leads back to the most basic question about utilitarianism: Is the greatest happiness principle the ultimate foundation of morality?

Chapter 4: Of what sort of Proof the Principle of Utility is Susceptible

Summary

Mill begins this chapter by saying that it is not possible to prove any first principles by reasoning. How, then, can we know that utility is a foundational principle? The purpose of this chapter is to explore what should be required of utilitarianism in order for it to be believed as valid. Mill argues that the only proof that something is desirable is that people actually desire it. It is a fact that happiness is a good, because all people desire their own happiness. Thus, it is clear that happiness is at least one end, and one criterion, of morality.

However, in order to show that happiness is the sole criterion for morality, it is necessary to show that people never desire anything but happiness. Mill says that people do desire things like virtue, which in common language is distinguished from happiness. However, Mill states that people love virtue only because it constitutes a part of happiness. Mill argues that happiness is not an abstract idea, but a whole with component parts. Because virtue is a part of happiness, and promotes the general happiness, utilitarianism encourages the development of virtue.

Anything that is desired beyond being a means to happiness is desired because it is part of happiness. Thus, Mill explains that proving utilitarianism is a psychological question. The real issue is whether it is true that people only desire things that are part of happiness or a means to happiness. This can only be answered by self-reflection and observation of others. Mill contends that utilitarianism is true, and that impartial reflection will show that desiring something is the same thing as thinking it pleasant. He argues that this is so obvious that he doubts it could be disputed. The only possible refutation that could legitimately be made is that the moral will is something different than physical or emotional desire; virtuous people carry out actions without thought of such pleasures. Mill admits that will is different than desire, and often becomes an end in itself. However, all will originates in desire; if we will a thing that we now no longer desire, it is only by force of habit. This does not change the fact that things are good to people only insofar as they lead to pleasure. Mill then says that it leaves it to the "thoughtful reader" whether what he has said is true.

Commentary

Mill further expands his discussion of happiness in this chapter. Recall that in Chapter 2, Mill argued that pleasures that were based on one's higher faculties were of a higher quality, and should be weighted accordingly. In this way, he tried to expand the meaning of happiness to allow for different kinds of pleasure. In Chapter 4 Mill expands the meaning of happiness again. A possible objection to utilitarianism is that certain experiences could be integral parts of a compound happiness, not merely a means to a pure, elemental happiness. Correspondingly, Mill argues now that utilitarianism can leave room for the fact that happiness consists of the other experiences that people value. This idea of happiness as having "component parts" is an important expansion of the meaning of happiness by Mill.

The other major argument in this chapter is that the motivation for all action is based on the fulfillment of desire. However, he probably rightly contends that whether he is correct is an empirical question, a question answered by observing oneself and others. This brings up an important question about the lines between psychology and philosophy. If utilitarianism is based on the psychological make-up of human beings, then to what degree is it merely descriptive? We tend to want philosophy to provide reasons why we should behave in a particular manner. However, to note that we do behave in a certain manner is not necessarily to prove that we ought to behave that way. One should consider at which points in the text Mill is observing how humans view the world, and at which points he is advocating a certain worldview. What

does his theory lose and gain from relying on psychological arguments? To what degree is it even possible to avoid a dependence upon description?

PART VI: POLITICAL PHILOSOPHY

ESSAY 24: A Theory of Justice, John Rawls

正义论

Justice as fairness 公平即正义

I. THE ROLE OF JUSTICE

A. JUSTICE AS THE FIRST VIRTUE: Rawls tells us that justice is the "first virtue" of social institutions. It is the standard by which they are to be judged, just like truth is the standard by which statements are to be judged.

B. ANTI-UTILITARIANISM: each person possesses an inviolability that we cannot ignore simply to maximize social welfare (= total utility). The way to ensure this is by taking equal liberties "as settled," i.e., as inalienable and non-negotiable.

C. SOCIETY AS A COOPERATIVE ENDEAVOR FOR MUTUAL BENEFIT: "There is an identity of interests in social cooperation;" that is, we all benefit from being in society (as opposed to being in a state of nature). Thus, being in society creates benefits.

D. THE ROLE OF PRINCIPLES OF JUSTICE: We need a set of principles for deciding how to distribute these benefits.

E. INTERLUDE: NOGGLE'S PARABLE OF THE ORCHARD.

F. A WELL-ORDERED SOCIETY: Definition of a well-ordered society as, roughly, a society in which people agree about what is just. This definition is important to Rawls's later work, but we probably won't get into it here. Because most societies are not well-ordered, they need some way to settle disagreements about what is just.

G. CONCEPT VERSUS CONCEPTION: The CONCEPT OF JUSTICE is: that benefits and burdens of cooperation are to be distributed properly. Various CONCEPTIONS OF JUSTICE are theories about what "proper" means here.

H. OTHER BASES FOR CHOICE OF SOCIAL INSTITUTIONS: This is a methodological point: while justice is the First Virtue of Social Institutions, we will not choose a theory of justice only on the basis of how fairly it distributes the benefits and burdens of cooperation.

We will also want to look at some of the broader consequences of accepting a particular theory of justice before we decide whether to accept it. This methodological point is crucial for understanding the New Rawls, but we won't worry too much about it.

II. THE SUBJECT OF JUSTICE

A. APPLICATION OF THE THEORY: Rawls's theory will apply primarily to the basic structure of society. That is, it will determine what the basic institutions should look like. For example: should we have private property, redistributive (progressive) income taxes, a capitalist economy, monogamous families, and so on. It will also tell us about what basic rights there should be. The theory will not "micro-manage;" that is, it will not tell us about specific laws, tax-rates, and so on.

B. AGNOSTIC ABOUT OTHER INSTITUTIONS: The theory is only meant to apply to the State. It may or may not work for private associations like clubs, families, etc.

C. STRICT COMPLIANCE THEORY: The theory is meant to apply only in societies characterized by "strict compliance." This is a somewhat utopian qualification which may (or may not) radically limit the application of the theory to actual societies.

D. HISTORICAL NOTES: The rest of the section (p. 9f) is more or less irrelevant for our purposes.

III. THE MAIN IDEA OF THE THEORY OF JUSTICE

A. SOCIAL CONTRACT THEORY: Rawls announces his allegiance to the social contract tradition: justice equals fairness, and fairness is to be explained in terms of what a person could agree to "in an initial situation of equality" if they were free and rational. Each part of this formulation is important, especially the part about equality.

B. THE ORIGINAL POSITION: This "original position" plays the same role in Rawls's theory of Justice as Fairness as the State of Nature plays in traditional social contract theory.

C. THE VEIL OF IGNORANCE: Description of the Veil of Ignorance (VI) and the Original Position (OP).

D. THE VEIL OF IGNORANCE, CHANCE, AND THE TWO MORAL POWERS: The initial situation behind the VI is "fair between individuals as moral persons, that is, as rational beings ...." Three things to note here:

1. The VI is what it is because it has to be that way to be fair. It deprives us of any knowledge of those things about us that are true because of chance. Thus it prevents us from making decisions based on "natural advantages" that we may have due to luck and social circumstances.
2. Rawls claims that, for his purposes, behind the veil of ignorance, people have two main features. He calls these the "two moral powers." The first is the ability to set and pursue their own ends, to construct and implement a "rational plan of life," and to adopt and be guided by a "conception of the good." The second is to have and be guided by a "sense of justice." That means that they can internalize moral norms and act upon them.
3. THE TWO MORAL POWERS ARE THE MAIN THINGS THAT ARE RELEVANT BEHIND THE VEIL OF IGNORANCE. From behind the veil, we know almost nothing about ourselves and others except that we have the two moral powers. We also know basic facts about the social sciences and economics. (More detail will be added in a later section.)

E. THE ORIGINAL POSITION AS A LOCATION FROM WHICH FAIR AGREEMENTS CAN BE MADE: The basic idea then, is that because the OP is fair, anything you would agree to in it is also fair. In other words, if you object to anything you would have agreed to in the OP, then it must be because you would rather have chosen in some way that you could not have behind the VI. To do that, you would have had to exploit knowledge about your particular situation that is ruled out by the VI. And to do that, Rawls wants to insist, would be unfair. The basic idea then, is that any society you would choose in order to gain an advantage from your social status, gender, race, natural talents, conception of the good (e.g. your religion) would be unfair. The VI keeps you from choosing a set of institutions that would discriminate on the basis of any of the things the VI makes you forget. In that way, it guarantees, says Rawls, that the outcome is fair.

F. THE FOUR STAGE SEQUENCE: The OP is meant to be a decision procedure for setting very basic principles of justice. That is stage one. The idea is that once these principles are chosen, then a constitution can be formed that will embody those principles. This is stage two. The constitution will then set up a legislature, which will enact laws that accord with the constitution and therefore the basic principles of justice. That is stage three. Then, judges and administrators will implement those laws. This is stage four. (More details about these stages can be found in section 31, which is optional reading for this course.)

G. THE MOTIVATIONS OF THE PARTIES: What motivates people in the OP? They are mutually disinterested but not necessarily egoists. They do not know what their goals are, but they do want to try to achieve them.

H. A PRELIMINARY ARGUMENT AGAINST UTILITARIANISM: If you do not know who you will be in society, you will not want to agree to a system that allows one person to be sacrificed for the good of others, since you might turn out to be that one person.

I. A PRELIMINARY ARGUMENT FOR RAWLS'S TWO PRINCIPLES OF JUSTICE: Instead, he thinks we would agree to two

main principles: basic equality of rights, and (what will later be called the Difference Principle) that social inequalities are only just if they help out the least fortunate.

J. PLAN OF THE OVERALL ARGUMENT: The argument has two parts: one is that the OP is a fair way to choose the principles to regulate the basic structure of society; the other is that the two principles he proposes would be chosen in the OP. These are separate arguments.

IV. THE ORIGINAL POSITION AND JUSTIFICATION

A. BASIC CLAIM: Rawls claims that what makes a political system just is that it would be chosen in the OP. This section will be an initial step in justifying that claim. It is not the whole argument, but rather a kind of "preview" of the argument that will unfold in later sections.

B. THE OP AS AN UNBIASED VANTAGE POINT: The OP is a way of making vivid the idea that societies should not advantage or disadvantage persons because of race, gender, social position, religion, intelligence, wealth, etc. To ensure this, we imagine what it would be like to choose a society if we did no know our race, gender, financial status, religion, etc. If we do not know these things, we cannot "rig" our decision about what kind of society to adopt in such a way as to guarantee our own advantage.

C. A DEVICE OF REPRESENTATION: "The Original Position is at Hand:" The OP is just a metaphor, a device to help us envision a society that set up fairly. Pretending to be in the OP is thus simply a tool to help us think about what kind of society is fair and just.

D. REFLECTIVE EQUILIBRIUM. This idea is very important in meta-ethics and normative epistemology. But we won't talk in much detail about it here. Reflective equilibrium is a theory about how to figure out what moral principles are the right ones. Basically the rough idea is this: we start out with the most obvious, unquestionable moral principles we can think of. We then produce a theory that makes the best sense of them. If that theory conflicts with other (less obvious, more questionable) moral beliefs, then we will either adjust the theory or change our beliefs. Once we get the theory and our beliefs to match up, then we are in "reflective equilibrium." This applies to Rawls's theory int eh following way: we start out with certain unquestionable moral principles, e.g., a person is not more valuable just because he is white or a male or a Protestant. We then construct a theory that will fit this belief. This theory is the OP. We can then use this new theory to help us answer questions that we don't know the answers to, such as: what should the basic features of society be like. As I say, this is more meta-ethics and moral epistemology than I want to get into in much detail. If you don't understand it, you can either see me about it during office ours, or simply ignore it.

E. MOTIVATION FOR THE OP AND ITS RESULTS: FAIRNESS: Why should we care what we would agree to in the OP? The answer should be fairly obvious if you've been reading along, but nevertheless, many people don't get it. The answer is that the OP embodies conditions that are fair. If we care about what is fair (as most of us do), then we should care what we would agree to in the OP, because what we would agree to there is fair. Or so Rawls argues.

V. CLASSICAL UTILITARIANISM

XI. TWO PRINCIPLES OF JUSTICE

A. RAWLS'S PRINCIPLES OF JUSTICE: In this section, Rawls is going to give us what he thinks is the answer to the question: "What would we choose in the OP?" A bit later he will give you a more detailed argument as to why we would choose these two principles in the OP.

B. AN INITIAL FORMULATION: The formulation of the Two Principles of Justice he gives here is not final. Throughout *A Theory of Justice*, he keeps fine tuning the formulation of the principles. Most of that tinkering will not concern us here. For the most part we can work with this initial statement (and I've simplified even more here).

C. THE TWO PRINCIPLES OF JUSTICE:

First Principle: Each person is to have an equal right to the most freedom compatible with everyone else having that same amount of freedom

Second Principle: any social and economic inequalities (=inequalities in the distribution of primary goods) MUST be arranged so that:

(a) they are "attached" to positions and offices open to all (this is often called "fair equality of opportunity.")

(b) they are to everyone's advantage

D. THE DEMOCRATIC INTERPRETATION OF THE SECOND PRINCIPLE: In sections 12-14, Rawls considers various possible ways to further refine the second principle. He considers several different ways to "interpret" the second principle. The one he settles on (for reasons that are more technical than I want to get into in an undergraduate course) is called "Democratic Equality." According to the democratic interpretation–which is the interpretation that Rawls settles on–the Second Principle should be read the following way:

"The higher expectations of those better situated are just if and only if they work as part of a scheme which improves the expectations of the least advantaged members of society. The intuitive idea is that the social order is not to establish and secure the more attractive prospects of those better off unless doing so is to the advantage of those less fortunate." (65).

This refined interpretation of the second principle is called the **Difference Principle**.

E. THE FINAL FORM OF THE TWO PRINCIPLES OF JUSTICE: For convenience, here is the full set of the Two Principles of Justice (2PJ) in their final form:

First Principle: Each person is to have an equal right to the most freedom compatible with everyone else having that same amount of freedom

Second Principle: any social and economic inequalities MUST be arranged so that they are:

(a) attached to positions and offices open to all ("fair equality of opportunity.")

(b) work to the benefit of the least advantaged group in society. (The **Difference Principle**)

F. IMPLICATIONS OF THE DIFFERENCE PRINCIPLE: The Difference Principle (DP) implies that it is OK for some people to have more money or social power IF allowing the them to have the extra power or wealth is good for everyone including the least fortunate, AND if each person has a fair chance of getting the extra power or wealth.

G. THE IDEA OF THE REPRESENTATIVE MAN: DEFINING SOCIAL POSITIONS. In section 16, Rawls considers several ways to define social groups. This is important, since the DP claims that inequalities of wealth and income can only be fair if they benefit the least advantaged group in society. Whether or not something is to the advantage of a group will, of course, depend on how exactly one defines who is in that group. Rawls looks at several possibilities. He does not settle on one particular one, but he does indicate that either of two general methods would be acceptable:

(a) The first method would be simply to identify a relevant kind of economic "player" or "social position", such as that of an unskilled worker.

(b) The second method would be to identify the least advantaged class simply by relative income. The suggestion Rawls likes is to define the lowest economic class as the one which makes less than half the median income in the society.

H. BASIC STRUCTURE AS SUBJECT: These principles apply only to the basic structure (that is, the main institutions) of society. They are not to be used to micro manage the economy, or tax rates, or anything like that.

I. SERIAL ORDER: This is sometimes called "lexical" order or "lexical priority." If we cannot satisfy both of the principles at the same time, then we MUST satisfy the first one first. (Later he introduces some exceptions to this, mainly for under-developed societies, but we will for the most part ignore the exceptions.)

J. CONSEQUENCES OF THE SERIAL ORDERING OF THE PRINCIPLES: No one can (either voluntarily or involuntarily) trade off her political freedoms for more power, wealth, etc. We make sure we all have equal basic freedoms and basic rights, and then we worry about dividing up the other stuff.

K. PRESUMPTION IN FAVOR OF EQUALITY: The Two Principles of Justice are a way to make specific a much more general conception of justice (which will sometimes be called the "general conception of justice"): all social values should be distributed equally unless an unequal distribution is to everyone's advantage.

L. PRIMARY GOODS: Rawls introduces this term to cover all the things that the Two Principles of Justice are going to divide up. They are things like rights, opportunities, incomes, power, and so on. More on this in the section 15.

XXVI. THE REASONING LEADING TO THE TWO PRINCIPLES OF JUSTICE

ESSAY 25: Anarchy State and Utopia, Robert Nozick

John Rawls and Robert Nozick: liberalism vs. libertarianism

John Rawls, "A Theory of Justice." Rawls' presents an account of justice in the form of two principles: (1) liberty principle= people's "equal basic liberties" — such as freedom of speech, freedom of conscience (religion), and the right to vote — should be maximized, and (2) difference principle= inequalities in social and economic goods are acceptable only if they promote the welfare of the "least advantaged" members of society. Rawls writes in the social contract tradition. He seeks to define equilibrium points that, when accumulated, form a civil system characterized by what he calls "justice as fairness." To get there he deploys an argument whereby people in an "original position" (state of nature), make decisions (legislate laws) behind a "veil of ignorance" (of their place in the society– rich or poor) using a reasoning technique he calls "reflective equilibrium." It goes something like: behind the veil of ignorance, with no knowledge of their own places in civil society, Rawls posits that reasonable people will default to social and economic positions that maximize the prospects for the worst off– feed and house the poor in case you happen to become one. It's much like the **prisoner's dilemma** in game theory. By his own words Rawls = "left-liberalism".

John Rawls, "A Theory of Justice." Rawls' presents an account of justice in the form of two principles: (1) liberty principle= people's "equal basic liberties" — such as freedom of speech, freedom of conscience (religion), and the right to vote — should be maximized, and (2) difference principle= inequalities in social and economic goods are acceptable only if they promote the welfare of the "least advantaged" members of society. Rawls writes in the social contract tradition. He seeks to define equilibrium points that, when accumulated, form a civil system characterized by what he calls "justice as fairness." To get there he deploys an argument whereby people in an "original position" (state of nature), make decisions (legislate laws) behind a "veil of ignorance" (of their place in the society– rich or poor) using a reasoning technique he calls "reflective equilibrium." It goes something like: behind the veil of ignorance, with no knowledge of their own places in civil society, Rawls posits that reasonable people will default to social and economic positions that maximize the prospects for the worst off– feed and house the poor in case you happen to become one. It's much like the **prisoner's dilemma** in game theory. By his own words Rawls = "left-liberalism".

Robert Nozick, "Anarchy, State, and Utopia," libertarian response to Rawls which argues that only a "minimal state" devoted to the en-

forcement of contracts and protecting people against crimes like assault, robbery, fraud can be morally justified. Nozick suggests that "the fundamental question of political philosophy" is not how government should be organized but "whether there should be any state at all," he is close to **John Locke** in that government is legitimate only to the degree that it promotes greater security for life, liberty, and property than would exist in a chaotic, pre-political "state of nature." Nozick concludes, however, that the need for security justifies only a minimal, or "night-watchman," state, since it cannot be demonstrated that citizens will attain any more security through extensive governmental intervention. (Nozick p.25-27)

" …the state may not use its coercive apparatus for the purpose of getting some citizens to aid others, or in order to prohibit activities to people for their *own* good or protection." (Nozick Preface p.ix)

Differences:

1. The primary difference between the two is in the treatment of the legitimacy of governmental redistribution of wealth (and even on that issue Nozick eventually flinches — see #1 below). In place of Rawls's "difference principle," Nozick espouses an "entitlement theory" of justice, according to which individual holdings of various social and economic goods are justified only if they derive from just acquisitions or (voluntary) transfers. No safety nets allowed (acquisitions from social programs are not just because they are funded through the involuntary transfer of wealth via taxation and are therefore taboo). No accommodations for free-riders should be made. Problem: Nozick never spells out the criteria of just acquisition.

- Nozick critique of Rawls's rationale for his difference principle: it's implausible to claim that merely because all members of a society benefit from social cooperation, the less-advantaged ones are automatically entitled to a share in the earnings of their more successful peers.

Similarities:

1. Both theories jump off with a sweeping statement of the primacy of justice — Nozick more or less retained Rawls's first principle (liberty) while rejecting the second (difference). But…
2. Regarding governmental redistribution of wealth, Nozick seems to admit that his entitlement theory is insufficient to refute demands for a redistributionist state; surely some collective holdings were acquired via some original act of unjust conquest, right?. In response Nozick agrees that a Rawls-like difference principle is morally acceptable after all, what he terms "rectification," on the premise that those currently least-well-off have the highest probability of being descended from previous victims of injustice. (Nozick p.152-153, 230-231)

3. Both shared a view of political philosophy as an exercise in the production of abstract theories, with little regard for the practical grounding of justice in human nature (i.e., of conformity with the likely demands of actual human beings). Therefore both theories rate a society's success by how closely it's laws and procedures adhere to the model rather than whether those laws produce morally maximized outcomes. Both clearly followed **Immanuel Kant**'s dictum, "let justice triumph, even if the world perishes by it."

Theory

Nozick's Entitlement Theory, influenced by John Locke, and Friedrich Hayek, which sees humans as ends in themselves and justifies redistribution of goods only on condition of consent, is a key aspect of *Anarchy, State, and Utopia*.

The book also contains a vigorous defense of minarchist libertarianism against more extreme views, such as anarcho-capitalism (in which there is *no* state and individuals must contract with private companies for *all* social services). Nozick argues that anarcho-capitalism would inevitably transform into a minarchist state, even without violating any of its own non-aggression principles, through the eventual emergence of a single locally dominant private defense and judicial agency that it is in everyone's interests to align with, because other agencies are unable to effectively compete against the advantages of the agency with majority coverage. Therefore, he felt that, even to the extent that the anarcho-capitalist theory is correct, it results in a single, private, protective agency which is itself a de facto "state." Thus anarcho-capitalism may only exist for a limited period before a minimalist state emerges.

Distributive justice

Nozick's discussion of Rawls's theory of justice raises the dialogue between libertarianism and liberalism to an epic level. The entitlement theory is sketched. In slogan form it states, "From each as they choose, to each as they are chosen". It comprises a theory of (1) justice in acquisition; (2) justice in rectification if (1) is violated (rectification which might require apparently redistributive measures); (3) justice in holdings, and (4) justice in transfer. Assuming justice in acquisition, entitlement to holdings is a function of repeated applications of (3) and (4). Nozick's entitlement theory is a non-patterned historical principle. Almost all other principles of distributive justice (egalitarianism, utilitarianism) are patterned principles of justice. Such principles follow the form, "to each according to..."

Nozick's famous Wilt Chamberlain argument is an attempt to show that patterned principles of just distribution are incompatible with liberty. He asks us to assume that the original distribution in society, D1, is ordered by our choice of patterned principle, for instance Rawls's Difference Principle. Wilt Chamberlain is an extremely popular basketball player in this society, and Nozick further assumes 1 million people are willing to freely give Wilt 25 cents each to watch him play basketball over the course of a season (we assume no other transactions occur). Wilt now has $250,000, a much larger sum than any of the other people in the society. The new distribution in society, call it D2, obviously is no longer ordered by our favored pattern that ordered D1. However Nozick argues that D2 is just. For if each agent freely exchanges some of his D1 share with WC and D1 was a just distribution (we know D1 was just, because it was ordered according to your favorite patterned principle of distribution), how can D2 fail to be a just distribution? Thus Nozick argues that what the Wilt Chamberlain example shows is that no patterned principle of just distribution will be compatible with liberty. In order to preserve the pattern, which arranged D1, the state will have to continually interfere with people's ability to freely exchange their D1 shares, for any exchange of D1 shares explicitly involves violating the pattern that originally ordered it.

Nozick analogizes taxation with forced labor, asking the reader to imagine a man who works longer to gain income to buy a movie ticket and a man who spends his extra time on leisure (for instance, watching the sunset). What, Nozick asks, is the difference between seizing the second man's leisure (which would be forced labor) and seizing the first man's goods? "Perhaps there is no difference in principle," Nozick concludes, and notes that the argument could be extended to taxation on other sources besides labor. "End-state and most patterned principles of distributive justice institute (partial) ownership by others of people and their actions and labor. These principles involve a shift from the classical liberals' notion of self ownership to a notion of (partial) property rights in *other* people."

Nozick then briefly considers Locke's theory of acquisition. After considering some preliminary objections, he "adds an additional bit of complexity" to the structure of the entitlement theory by refining Locke's proviso that "enough and as good" must be left in common for others by one's taking property in an unowned object. Nozick favors a "Lockean" proviso that forbids appropriation when the position of others is thereby worsened. For instance, appropriating the only water hole in a desert and charging monopoly prices would not be legitimate. But in line with his endorsement of the historical principle, this argument does not apply to the medical researcher who discovers a cure for a disease and sells for whatever price he will. Nor does Nozick provide any means or theory whereby abuses of appropriation—acquisition of property when there is *not* enough and as good in common for others—should be corrected.

APPENDIX: SOME PHILOSOPHICAL TERMS

ENTAILMENT蕴涵: A conclusion is entailed by its premises前提 just in case it's impossible for all the premises to be true and the conclusion false.

entailment is the relationship between two sentences where the truth of one (A) requires the truth of the other (B).

AN ARGUMENT IS **VALID**有效 JUST IN CASE ITS CONCLUSION IS ENTAILED BY ITS PREMISES.

AN ARGUMENT IS **SOUND**合理 JUST IN CASE IT'S VALID AND ALL ITS PREMISES ARE TRUE.

MODUS PONENS:肯定前件 if P then Q, P => Q

MODUS TOLLENS:否定后件 if P then Q, not Q => not P

DISJUNCTIVE SYLLOGISM:选言三段论 Either P or Q, not P => Q

CATEGORICAL SYLLOGISM:直言三段论 All Fs are G, x is an F => x is G or if given

FALLACY OF EQUIVOCATION:一词多义

Sometimes it's hard to come up with a direct prove of "p", instead, you can assume not("p"), then if you can get something like both "q" and "not q", namely a CONTRADICTION, then you proved "p" now. The method is called **REDUCTIO AD ABSURDUM (**反证法**).**

RATIONAL THEISM: there is a rational basis for belief in god.

ARATIONAL THEISM: there's no rational basis for belief in god, but one should believe in god anyway.

IRRATIONAL THEISM: there's a rational basis for believing that god DOESN'T exist, but one should believe in god anyway.

ATHEISM: GOD DOESN'T EXIST. 无神论

AGNOSTICISM: WE CAN'T KNOW WHETHER GOD EXISTS OR NOT. 不可知论

CAUSAL DETERMINISM (CD): A statement of all the facts at time t (inc. facts about people's heredity, environment etc.), together with a statement of the law's of nature, logically entails what will happen at any future time $t + 1$.

INCOMPATIBILISM: only one of CD and human freedom is true.

hard determinism: a broad view that causal determinism is incompatible with human freedom, and CD is true, so human freedom must be false, it's an illusion.

libertarianism (Chisholm): it agrees with hard determinsm if causal determinism were true that human freedom is false, but what they differ is that libertarianism believes that we are free, so it is that CD is false.

COMPATIBILISM: both of CD and human freedom are true, they accept 1 is true, gonna choose either 2 or 3 true) (Ayer chose 3 true.

Numerical identity: X and Y are numerically identical if and only if they are one and the same thing. (e.g. Bruce Wayne and Batman)

Qualitative identity: X and Y are qualitatively identical if and only if they have exactly the same properties. (e.g. identical watches, Big Macs)

Ability Knowledge e.g. I know how to ride a bike

Factual Knowledge e.g. I know that the giants won the world series last year.

Acquaintance Knowledge e.g. I know Rob Ford.

"JTB = Justified True Belief" Theory of Knowledge: P stands for a statement / proposition. S stands for "somebody".

S knows that P iff:

1. P is **T**rue (Knowledge is "factive")
2. S **B**elieves that P (Psychological attitude)
3. S is **J**ustified in believing that P (Knowledge requires "having good reasons" for one's belief)

Skepticism: The view that we lack knowledge, that we don't know things we may **think** we know.

Psychological Certainty: One is psychologically certain of proposition P iff one has no doubts regarding P.

Evidential Certainty: One is evidentially certain of proposition P iff one's evidence for P "guarantees" P, that is, iff one's evidence is logically incompatible without not-P.

Skeptical Hypothesis A hypothesis that: i) Renders false most (or all) of our beliefs about the external world, and yet ii) We can't rule it out.

Psychological Egoism: the only thing anyone does pursue *as an end of itself* is her own self-interest.

Hedonism (Hedonistic Theory): Well-being consists in enjoying pleasant experiences and avoiding painful ones.

Desire Theory (Desire-Fulfilment Theory): Well-being consists in the satisfaction of desires

Objective List Theory: Well-being consists in living a life featuring items drawn from a list of things that are good in themselves.

Less Plausible reading: X's doing A has moral worth iff X ought do A, and X does A only "from duty" (i.e. is motivated to do A by and only by the thought that she ought do A.)

More Plausible reading: X's doing A has moral worth iff X ought to do A, and X is motivated to do A by the thought that he ought to do

A and whatever other motives she may also have, X would do A even if she had not had them.

Categorial Imperative:

1. "Act only on that maxim through which you can at the same time will that it become a universal law."
2. "Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as n end."
3. "Act always so as to aid in bringing about the kingdom of ends."

Utilitarianism

(1) You ought to do act A iff the happiness/unhappiness balance that would result from your doing A is **greater than** what would result from your not doing A;

(2) You ought **not** to act A iff the happiness/unhappiness balance that would result from your doing A is **less than** what would result from your not doing A.

2 principles of Justice

Principle 1 (p302-303): Each person is to have an equal right tot the most extensive scheme of "equal basic liberties" compatible with a similar scheme of liberties for others. *e.g.* Right to vote, freedom of speech, right to hold personal property etc…

Principle 2: aka the "**Difference Principle**": Social and economic inequalities are to be arranged so that they are both:

1. reasonably expected to be to everyone's advantage, and
2. attached to positions and offices open to all.

Principle 1 takes priority over Principle 2.