PH1102E

Week 2

Freedom and responsibility

I. Key concepts

II. Moral nihilism

III. Hume's alternative to nihilism

IV. An objection to Hume, and a reply on Hume's behalf

V. Concluding remarks

Strawson: We are not morally responsible for what we do, and thus we do not deserve any praise or punishment for what we do, but punishment is justified to deter people from doing the wrong things
Hume: The intent is all it matters, whether the factors are out of our control or not.

I. Key concepts.

Moral responsibility

You are morally responsible for doing X if by doing X, you give others a good reason for thinking well or badly of you. (Notice the word "good." Even if the Gestapo officer thinks well of Franz for turning over the Jewish family, he has no *good* reason to think well of Franz: his reason is the bad one that Franz's action facilitates the Nazi program of genocide.)

We can also say that someone has moral responsibility for those of his actions for which he deserves praise or blame, or reward or punishment.

Here, it is important to be aware of the fact that we can have justification for punishing (or rewarding) a person for causing some harm, even if the person does not deserve the punishment. For example, if a person suffering from a severe mental illness poses a threat to the community, the community is justified in locking him up, even though it's no fault of the demented person that he suffers from the mental illness that afflicts him. In this case, we are punishing someone (restricting his liberties), and we are justified in doing so, even though the person we're punishing does not *deserve* the punishment.

Similarly, when we cage a dangerous animal, we are not giving the animal what it deserves. The animal does not *deserve* to be caged; it would be absurd to point to the caged wildcat and say, "Well, he got what he deserved." But that doesn't change the fact that it's a good idea to cage the wildcat, if the cat poses a threat to the community. This is another case in which we are justified in restricting an agent's liberties, despite the fact that the agent does not deserve any such restriction.

In fact, we can even be justified in punishing someone for something the person didn't do! After all, the main rationale for punishing people is to deter them (and others) from engaging in harmful behavior. If everyone thinks that I murdered the old lady, and the judge sentences me to death for murdering her, this will tend to discourage other people from murdering old ladies, *even if I am actually innocent*. The judge's sentence has its deterrent effect regardless of whether I actually committed the crime.

Similar points apply to the positive side of moral responsibility. Suppose I spend all my free time doing volunteer work at a nursing home for the elderly. People might praise me for this; they might even give me a public service award, with the hope of inspiring other people to follow my example. But this rationale for rewarding me exists even if I do not deserve any reward -- even if, for example, I detest the elderly, and volunteer at the nursing home only to befriend rich old people in the hope that they will leave me money when they die.

So it's one thing to say that an agent deserves some punishment or reward, and another to say that there's a good rationale for punishing or rewarding the agent. We can have a good rationale for punishing or rewarding someone who does not deserve to be rewarded or punished. When it comes to moral responsibility, the question is whether we ever deserve praise, blame, reward, or punishment for our deeds, not whether it ever makes sense (for the sake of public safety, or maintaining social order, or whatever) to praise, blame, reward, or punish us for what we do.

Determinism

Determinism says that every event has a cause (except for the very first event, if there was a first event). Whether determinism is *true* is controversial; some think we live in a deterministic world, others disagree.

Indeterminism is simply the negation of determinism: it is the view that not every event has a cause; i.e., it says that some events (and not just the very first event, if there was a first event) do not have causes.

To say that one event, A, causes another event, B, is just to say that the occurrence of A makes the occurrence of B inevitable, given the laws of nature. For example, pouring water on a campfire causes the fire to go out, since the laws of nature dictate that if you saturate a burning object with water, the object stops burning.

Determinism says that every event is an inevitable result of earlier events. Here we're using the word "event" broadly, to include anything that happens or takes place in any way. When the Sun rises, that is an event. When a cell divides, that is another event. When a stock market crashes, that is yet another event.

Also included among events are the events that occur in your life, including the events that, so to speak, make up your biography. Your actions are events; a complete list of the events that take place in the world will include your actions.

This means that if determinism is true, each of your actions has a cause. But the causes of your actions are also events, and, therefore -- according to determinism -- also have causes. These causes (of the causes of your actions) also have causes, according to determinism, and so on back in time forever (or until we reach the very first event that ever took place, if there was a first event).

So, if determinism is true, each of your actions is an inevitable result of earlier events, where these earlier events resulted inevitably from yet earlier events, which resulted inevitably from yet earlier events, etc., etc., etc.

Now, if event C is an inevitable result of event B, and event B is an inevitable result of event A, then event C is an inevitable (albeit indirect) result of event A. So, if determinism is true, then each of your actions is an inevitable result of events that took place before you were even born. Events that occurred long ago set in motion a whole sequence of events that would eventually, and inevitably, result in your doing whatever it is that you do.

Determinism does not imply that you have no choice in what you do. Usually, when you do something, you do it because you want or choose to do it. You have a desire for coffee, and a belief that the nearest source of coffee is the Arts Canteen, and this belief and desire jointly cause you to walk to the Arts Canteen. All of this is perfectly compatible with determinism. It's just that determinism says that your belief and desire themselves have causes that preceded them in time, which in turn had even earlier causes, etc. Determinism does say that events in the distant past made it inevitable that you would walk to the canteen. But those distant events made it inevitable that you would walk to the canteen by

making it inevitable that you would **want** to walk to the canteen. Your beliefs, desires, choices, deliberations etc. are necessary links in the causal chain leading from the distant past to your present action. If you hadn't wanted coffee, or hadn't believed that there was coffee at the canteen, you wouldn't have acted as you did. It's just that, if determinism is true, it was already determined in advance that you would have the precise beliefs, desires, etc. that you did, in fact, have.

Determinism is therefore not the same as fatalism. According to fatalism, your future behavior is decided by pre-existing factors that will make you do the things you will do *regardless of whether you want to do them*. Determinism says that a thousand years ago, it was already inevitable that you would enroll in PH1102E, because it was already inevitable that you would desire to enroll in PH1102E, and already inevitable that nothing would prevent you from acting on that desire. By contrast, fatalism says that a thousand years ago, it was already inevitable that you would enroll in PH1102E even if you had no desire to do so, and even if you had a strong desire not to enroll, and even if I had no desire to teach the module!

Determinism is controversial, but may be true. Fatalism is just silly (outside the context of Greek tragedy anyway), and we won't talk about it any more.

II. Strawson's moral nihilism.

Galen Strawson thinks that it is impossible for anyone to bear moral responsibility for anything that he or she does. He thinks that the idea that people sometimes "deserve" to be praised or blamed (or punished or rewarded) for their actions is, ultimately, incoherent. He thinks that nothing anyone does ever gives us a *good* reason to think well or badly of the person. In a word: he thinks that there is no such thing as moral responsibility, and, therefore, no such thing as morality (right and wrong). This point of view is called **moral nihilism**.

Strawson's basic argument for moral nihilism is simple: everything you do results from factors over which you have no control; but you are not morally responsible for the results of such factors; therefore, you are not morally responsible for anything you do. To make it easier to discuss this argument, let's write it down as a sequence of steps, giving each step a separate name:

Step 1. The No Control Thesis:
Everything you do results from factors over which you have no control.

Conclusion: Whether determinism is true or not, we have no control over the factors, thus are not held morally responsible

Step 2. The Incompatibility Thesis:
You are not morally responsible for the results of such factors.

Conclusion: What we do are inevitable consequences of the factors that we have no control over, thus we are not morally responsible.

The conclusion (Moral Nihilism):
Therefore, you are not morally responsible for anything you do.

This is Strawson's basic argument for moral nihilism. Is the argument convincing? That depends on how convincing we find the steps leading up to its conclusion. If we follow Strawson in accepting statements 1 and 2 (the No Control Thesis and the Incompatibilist Thesis) we have no choice but to follow him in accepting the nihilistic conclusion which those statements imply. But why should we accept (1) and (2)?

The No Control Thesis

First, let's see why Strawson thinks we have to agree with statement (1). This is the statement that everything we do results from factors that we have no control over. To prove that this is true, Strawson considers two possibilities: the possibility that determinism is true, and the possibility that determinism is false (i.e., the possibility that indeterminism is true).

Suppose that determinism is true. Then, as pointed out earlier, every time you do something, your action is an inevitable (albeit indirect) result of events that took place long before you were born. Suppose you went clubbing last night instead of staying home to read philosophy. Then it was already decided in advance that you would go clubbing: given the laws of nature, and given the way the universe was configured a million years ago, there was no chance at all that you would stay home to read philosophy. In theory, someone living a million years ago could have predicted with absolute confidence that you would go clubbing rather than read philosophy on Saturday, August 20th 2011. At least, this is true if our world is a deterministic one.

Now, obviously, you have no control over what happened before you were born. You did not choose your parents, much less your parents' parents, much less the state of the cosmos five seconds after the Big Bang. Yet, according to determinism, the state of the cosmos five seconds after the Big Bang already made it inevitable that you would go clubbing some 12 billion years later. So, if determinism is true, this action, as well as every other action that you ever perform, is an inevitable result of factors that are completely beyond your control.

What can Strawson conclude at this point? Has he established that statement (1) is true? No, not yet. So far, he has only shown that the No Control Thesis is true if determinism is true. But maybe determinism is false. And if it is false -- if we live in an indeterministic universe, rather than a deterministic one -- then maybe some of our actions are not inevitable results of factors that we have no control over.

This is the idea behind Jean-Paul Sartre's philosophy of freedom and responsibility. According to Sartre, we are morally responsible for many of our actions, precisely because we often act upon conscious choices or desires that are not caused by anything else. For example, when I overcome the temptation to gain some advantage for myself by dishonest means, my mental choice to do the right thing is, in Sartre's view, an event that is free from any causal determination. When I make the choice, it is as if a miniature Big Bang takes place in me: the choice I make initiates a completely new chain of events, causally unconnected with anything that came before. It is this freedom of my inner, conscious choices from external causal constraint that makes me responsible for my actions -- or, for those of my actions that arise from these inner choices. That is Sartre's view.

Strawson thinks that Sarte is mistaken. Strawson thinks that even if our actions are not made inevitable by past events, we still are not morally responsible for performing them. Why not? Well, what does it mean to say that an event is causally undetermined? It means, among other things, that there is no reason why it takes place. If there were a reason why the event occurred, then presumably we could point to that reason as the cause of the event. An undetermined, uncaused event is by its very nature an event over which no one has any control. After all, if it's up to me whether or not the event takes place,

then the event is not causally undetermined: it is caused to occur (or, as the case may be, not occur) by me, or by my decision to let the event take place (or prevent it from taking place).

Strawson's point is that you have no more control over uncaused events than you have over events that took place before you were born. So even if Sartre is correct when he says that our actions arise from uncaused events, that doesn't change the fact that our actions arise from events over which we have no control.

So, putting it all together, Strawson reasons as follows. If determinism is true, then our actions all result from factors over which we have no control. Likewise, if indeterminism is true, our actions all result from factors over which we have no control. But either determinism or indeterminism has to be true: there is no third option. So we must conclude that our actions all result from factors over which we have no control.

Our actions result from factors over which we have no control, regardless of whether we live in a deterministic world. This is the No Control Thesis.

The Incompatibility Thesis

So far we have been considering Strawson's argument for the No Control Thesis. But in order to reach his ultimate objective, which is to prove moral nihilism, he needs another thesis: what I am calling the Incompatibility Thesis. This is the thesis (statement, claim, contention -- whatever you want to call it) that we're not morally responsible for anything that results from factors over which we have no control.

Why do I call this the "Incompatibility Thesis"? I call it this, because it says that the idea that you are morally responsible for doing X is incompatible with (conflicts with) the idea that you did X as a result of factors beyond your control (events that occurred before you were born, or randomly occurring events).

What is Strawson's argument for the Incompatibility Thesis? Why should we accept this part of his argument for moral nihilism?

Strawson is not very explicit about his reasons for taking the Incompatibility Thesis to be true, but I take it that the basic idea is this. Suppose someone deliberately does something to harm you. In this situation, your natural reaction will be to blame or punish the person for what he has done to you, or at least to think of him as deserving to be thought of in some negative way. But the more you think of this person and his behavior as resulting from factors entirely out of his control, the less you will be inclined to react to him in this way, and the more you will be inclined to think of him and his harmful behavior as a problem to be solved. In Strawson's view, the more we see people as products of forces beyond their control, the less we think of them as deserving any praise or blame for their behavior, and the more we see them as natural goods or natural evils, like sunshine and earthquakes.

If a snake bites me, I don't hold the snake morally responsible for the harm it has done. As Strawson sees it, this is because I recognize that the snake ultimately has no control over its behavior: it bites me because it is determined to do so by factors out of its control (its hard-wired instincts). But in this respect, the snake is no different from the rain that gets me wet, and I am no different from the snake

or the rain. We all do what we do as an inevitable consequence of factors that are completely beyond our control. If we are not going to hold the snake or the rain morally responsible for the harm they do, then consistency demands that we not hold human beings morally responsible for the harm they do either. This, at any rate, is Strawson's position.

By rejecting the concept of moral responsibility, Strawson rejects the suggestion that people ever deserve to be praised or blamed, or punished or rewarded, for what they do. But, as noted at the outset, this is consistent with saying, even insisting, that praising, blaming, punishing, and rewarding people is often a very sensible thing to do. It makes sense to punish rapists and murderers, because punishing them deters other people from raping and murdering. Strawson completely agrees with this. But according to Strawson, even though we are fully justified in punishing murderers, it doesn't follow that the murderers deserve punishment. According to him they do not deserve punishment, any more than deadly diseases deserve punishment. Murderers, like deadly diseases, are to be thought of as public health-hazards to be contained or eliminated, not as objects of righteous indignation or moral outrage.

III. Hume's alternative to nihilism.

According to Strawson, you are not morally responsible for anything that happens (or anything that you do) as a result of factors that are beyond your control. This is the second step of his argument for moral nihilism (the step that I have been calling the Incompatibility Thesis). Various philosophers have questioned the Incompatibility Thesis, the most famous example being David Hume.

According to Hume, you can be morally responsible for doing something , even if you are predetermined to do it by factors that are beyond your control. In Hume's view, you are morally responsible for what you do whenever you intentionally cause some good or harm---that is, whenever you bring about some beneficial or harmful state of affairs by acting on an intention to bring about that state of affairs.

Hume recognizes that not just any intention- or choice-driven behavior with a good outcome merits praise, and not just any intention- or choice-driven behavior with a bad outcome deserves blame. The intention that lies behind the action must be an intention to do harm (or good).

A bird is not like this. The bird has a desire to eat your fruit, and by eating your fruit it does you some harm; but there are no hurtful intentions behind the bird's action. The bird does not intend to do you any harm. Unlike a normal human being, a bird does not conceive of other creatures as having interests of their own which the bird can promote or interfere with. Likewise a shark that attacks a surfer does not conceive of what it is doing as harming the surfer; rather, the shark conceives of what it's doing as filling its belly. It is because birds and sharks do not intend to benefit or harm other creatures that birds and sharks have no sense of shame, or pride. When someone acts on a desire to harm, he does something that viruses, birds, sharks, etc. are not capable of. Human beings can conceptualize their behavior in a way that these other beings cannot.

Of course, we do not always hold human beings morally responsible for the harm (or good) that they do. But if we consider the cases in which we are not inclined to hold someone responsible for his harmful or

beneficial behavior, we find that these are cases in which the person did not act with an intention to cause benefit or harm. For example, I might kill you by giving you a peanut butter sandwich, not being aware of the fact that you have a deadly allergy to peanuts (maybe you yourself do not know that you have this allergy). In this case, we do not hold me responsible for the harm I've done, because I had no desire to harm you at all. We also tend not to hold very young children morally responsible for the harm they do, again on the grounds that they have not yet learned to see other people as having interests on a par with the child's own; small children are similar to birds, in this respect.

We know how Strawson will respond to all of this. He'll point out that when a person forms a good or bad intention, he is already performing a kind of action: an inner action of intention-formation. He'll also point out that, by Hume's account, the person is morally responsible for this act of intention-formation only if he performs it intentionally. I am morally responsible for intending to kill the old woman only if I intend to have this intention. And I am not morally responsible for intending to intend to kill the old woman unless I form the intention to intend intentionally, which I am morally responsible for doing only if my intention to form the intention to intend to kill the old woman is itself intentional...etc., *ad infinitum*.

Hume will agree with all of this. He will agree that I am not morally responsible for forming my various intentions (at least, not typically). But Hume will argue that this is irrelevant to the question of whether I am morally responsible for my actions. When it comes to the question of whether I am morally responsible for performing some overt action (killing an old lady, rescuing a drowning child, or whatever), the only relevant consideration is whether I performed the action from an intention to do harm (or good). The fact that the action and the intention behind it have their ultimate origins in factors over which I have no control is neither here nor there. Or, so says Hume.

Ultimately, the disagreement between Strawson and Hume is this. Strawson thinks that you are absolved of moral responsibility for doing X if you do X as a result of factors that are beyond your control (events in the distant past, randomly occurring events, or whatever). Hume thinks that you are not absolved of moral responsibility for doing X by the fact that you do X as a result of factors that are beyond your control, provided that you also do X as a result of your own good or bad intentions.

Strawson says: acting on a suitable intention is at most necessary, but not sufficient, for moral responsibility. Hume says: acting a suitable intention is certainly sufficient (even if not necessary) for moral responsibility. This is the fundamental disagreement between Strawsonian moral nihilists and Humean moral realists.

IV. An objection to Hume, and a reply on Hume's behalf.

I do not know who is right, Strawson or Hume. But there is one seemingly powerful objection to Hume that we should consider, if only to see that it may not, actually, be as powerful as it first appears.

Consider the case of Mr. Smith, a retired banker who has been happily married for 40 years to his wife, Mrs. Smith. Mr. Smith is just an ordinary decent sort of person, not without his flaws and weaknesses,

but no moreso than any other upstanding member of society. He has no dark secrets, violent fantasies, or suppressed neuroses.

Now, suppose that one day a mad scientist somehow secretly installs remote-controlled electrodes into Smith's brain. (Neither Smith nor anyone else is aware of this.) These electrodes allow the mad scientist to stimulate Smith's brain in such a way as to cause Smith to have any desire that he (the mad scientist) wants to Smith to have. It's not that the mad scientist has any direct control over Smith's muscles: the scientist cannot manipulate Smith's body like a marionette. All he can do with his remote-controlled electrodes is to cause various desires and other psychological states to arise in Smith.

Suppose the mad scientist now uses his electrodes to give Mr. Smith a desire to kill his wife. What will happen?

Suppose that Mr. Smith acts on his new desire and kills his wife. Should we hold him morally responsible in this case? It seems that we should. After all, Mr. Smith is not a mere animal: he ought to have suppressed the sudden desire to kill his wife. And given that he is, as we have supposed, a basically decent person who loves his wife, he would, in fact, suppress the murderous desire that the mad scientist stimulated in him.

But what if we take the case a step farther? Suppose that the scientist not only gives Mr. Smith a desire to kill his wife, but simultaneously suppresses Mr. Smith's psychological desire-suppression mechanisms. Suppose that the mad scientist deletes Mr. Smith's happy memories of his life with Mrs. Smith, and replaces them with false memories of a strife-ridden marriage to an insufferable shrew of a wife. Suppose that the scientist alters Mr. Smith's psychological makeup in whatever way he must, in order to get Smith to act on the intention (also implanted by the mad scientist) to kill his wife.

If the mad scientist does all this, then Mrs. Smith is in trouble. With no psychological brakes or circuit-breakers to prevent him from acting on his implanted intention to kill his wife, it looks as though Mr. Smith is doomed to go through with it.

Suppose he does go through with it: he stabs his wife forty-two times with kitchen knife. What are we to say? Are we to say that Mr. Smith is morally responsible for killing his wife? Surely not. Surely Mr. Smith's role in this story is purely that of a victim, not that of a culprit. The mad scientist was the murderer; Mr. Smith was merely the murder weapon.

Yet it seems as though Hume must say that Mr. Smith <u>is</u> morally responsible for killing his wife. After all, in killing her, he acted on an intention to harm her. True, his intention arose from factors beyond his control (namely, the nefarious meddling of the mad scientist). But according to Hume, all that is required for moral responsibility is that you act on an intention to cause some benefit or harm. The fact that your intention is an inevitable consequence of some other factors that are beyond your control is irrelevant. The fact that Smith is compelled to form his murderous intention by the mad scientist therefore does not absolve Smith of moral responsibility, according to Hume's theory. But since Smith is clearly <u>not</u> morally responsible for killing his wife in this scenario, Hume's theory must be wrong.

How should Hume reply to this? I suggest that his best reply is to agree that Mr. Smith is not morally responsible for killing his wife, but to insist that the person who stabbed Mrs. Smith is morally responsible for her death. In other words, Hume can insist that the person who stabbed Mrs. Smith is indeed morally responsible for killing her, but deny that it was Mr. Smith who killed Mrs. Smith. The mad scientist destroyed Mr. Smith when he altered his psychology in the ways described, replacing him with a new (and dangerous) person. In effect, the mad scientist killed Mr. Smith, and planted a new person in Mr. Smith's body. It is this new person, not Mr. Smith, who bears moral responsibility for Mrs. Smith's death. And this person did act on a harmful (indeed murderous) intention. Thus the example does not constitute a counterexample to Hume's theory of moral responsibility.

How effective is this defense of the Humean theory? In the end, it comes down to a question of the self and personal identity, which we'll address in Week 10. According to some theories of the self, which focus on psychological criteria of personal identity, Hume is right to insist that the mad scientist's interference results in the destruction of Smith (and the creation of a new person). According to other theories, Smith survives, and therefore ought, by Hume's logic, to be held responsible for the death of his wife.

V. Concluding remarks.

Who is right, Strawson or Hume? As I said before, I'm not sure. At the end of the day, it seems to come down to the question of which we should give greater importance to: the fact that all our behavior arises from factors that are beyond our control (events in the distant past, random occurrences, etc.), or the fact that some of our behavior also arises from our own intentions (desires, decisions, choices) to bring about good or bad results. Perhaps the most surprising outcome of these deliberations is that what is in many ways the most natural conception of moral agency---the Sartrean conception---proves to be the least tenable. However we solve the problem of freedom and responsibility, whether *à la* Hume or *à la* Strawson, the solution is likely to come at some cost to naive common sense.