Class 14 - Free Will I

Woolfolk, Doris, and Darley, "Identification, Situation Constraint, and Social Cognition: Studies in Attribution of Moral Responsibility"

I. The Classic Free Will/Determinism Debate

The basic problem of free will and determinism comes in two versions, theistic and physical.
In both versions, we start with the basic intuition of free will.
It seems that I am free to decide, within the laws of physics, what I am going to do.
I am free to continue writing this sentence, or to dance a jig outside in the cold rain.
Indeed, this freedom seems to be essential to our senses of who we are.
We identify ourselves, to a large degree, with the free choices we make.
From the inside, we are free.

In contrast, there are two ways in which my apparently free choices appear, from the outside, to be determined.
In the theistic version, we imagine that there is an omniscient, omnipresent being.
Such a being, let's call her Alice, knows what I am going to do at all times, including future times.
So, Alice knows that I will continue typing this sentence.
Alice knows what I will wear tomorrow, and next week, and at all future times.
Since Alice knows, even now, every choice I will make in the future, it seems that my choice is already determined.
This predetermination conflicts with the intuition I have that I am free to choose.

In the physical, or Laplacian, version of determinism, we imagine that our world is completely governed by deterministic physical laws.
Most physical laws seem deterministic in that they yield specific and determinable results, given a complete description of initial conditions.
Just as the apple has no choice but to fall from the tree, my actions are governed by laws of neuroscience and physics generally.
The future appears to be fixed, beyond our control.

One response to the determinist is to show that the future is not fixed.
We might do that by appealing, say, to the indeterminacy of quantum physics.
But, quantum indeterminacy does not seem to rise to the observable level.
Moreover, the deterministic-seeming laws of physics do not suffer from random indeterminacies.
Indeed, if they did, not only would they seem undetermined; they would seem chaotic.
Our freedom does not seem to consist of random moments inconsistent with the laws.
Our freedom is rooted in our ability to choose among various options.

On the physical version of determinism, any appearance of free will can be attributed to a lack of understanding of the laws and the initial conditions.
If my mind, for example, is just my brain, and my brain works according to strict physical laws, then my thoughts, as well as my overt behavior, are determined.

Determinism, especially physical determinism, seems troubling for a variety of reasons.
First, there is just the unpleasant thought that I don't have the freedom I appear to have.

Less phenomenally, determinism seems to undermine our notions of moral responsibility.
Ordinarily, we think that we are morally responsible only for behavior that we could have avoided; we are not responsible when we have no ability to do otherwise.
So, I am not personally responsible for, say, ending climate change, since I can not personally stop it.
I am not responsible for, say, tidying the surface of Jupiter or preventing the great Chicago fire of 1871 since the laws of physics prevent me from having any effect on distant objects or past events.
On the other hand, since I can contribute, in some way, to reducing our carbon output, say by sometimes walking to work instead of driving, I may be responsible for doing so.
But, if determinism is true, and if it entails that I can never do otherwise than what I do, it seems that I can never be morally responsible for any of my actions.
Intuitively, we do think people are morally responsible for some of their actions.
So, determinism clashes with these intuitions.
This is a puzzle.

II. Compatibilism

A standard way of trying to resolve the puzzle is to adopt a compatibilist notion of free will.
Compatibilism, which Hume defended, is the view that, contrary to the incompatibilist view described above, determinism is not opposed to free will.
According to Hume, an act is free if it is done in accordance with our will.
If I do something only because I could not have done otherwise, I do not do it freely.
I do not return to the ground when I jump in the air of my free will; I could not have done otherwise.
More importantly, if I pay my taxes because I am afraid of being fined or imprisoned, or if I refrain from cheating only out of fear of punishment, or if I am forced by threat to do any action I do not wish to perform, I do not act freely.
On the other hand, if I want to pay taxes, since I approve of their uses in building and maintaining roads, schools and armed forces; or if I refrain from cheating because I do not wish to cheat, then I am acting in accordance with my will, freely.
Consequently, we can hold people morally responsible for those acts they perform freely, in Hume's sense, and not for those they perform under constraint.

Hume's notion of free will allows us an account of moral responsibility which aligns with our belief that we are responsible only for that which we choose.
Hume's definition is consistent with the doctrine that ought implies can, that our moral responsibilities do not exceed our powers.

But, determinists tend to be unsatisfied with Hume's definition of freedom, because it fails to take into account any constraints on our will.
The determinist pursues the question of whether we are free or determined by asking whether we are free to choose what we choose or whether we are constrained.
If our thoughts are themselves the products of physical processes, mainly brain processes along with their inputs (from perception), then the same problem of determinism recurs with regard to our will.
Our actions may be in accord with our will, but we are prevented from willing freely.
The determinist who digs-in his/her heels against the compatibilist is called an incompatibilist: free will and moral responsibility are incompatible with determinism.

III. Incompatibilism

The most influential argument for incompatibilism is called the consequence argument.
Here is a version of the consequence argument from Peter van Inwagen:

> If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. ("An Essay on Free Will," p 56)

Here is another version, from David Lewis.

> Suppose that determinism is true, and that I just put my hand down on my desk. As a compatibilist, I claim that this is a free but determined act. I was able to act otherwise, for instance to raise my hand. But there is a true historical proposition H about the intrinsic state of the world long ago, and a true proposition L specifying the laws of nature, such that H and L jointly determine what I did, and jointly contradict the proposition that I raised my hand. If I had raised my hand, then at least one of three things would have been true: contradictions would have been true, H would not have been true, or L would not have been true. So if I claim that I am able to raise my hand, I am committed to the claim that I have one of three incredible abilities: the ability to make contradictions true, the ability to change the past, or the ability to break (or change) the laws. It's absurd to suppose that I have any of these abilities. Therefore, by reductio, I could not have raised my hand (paraphrased from "Are We Free To Break the Laws" by Kadri Vihvelin, in http://plato.stanford.edu/entries/incompatibilism-arguments/).

The central claim of the incompatibilist returns us to the Laplacian puzzle.
We can not choose among the consequences of our actions.
Thus, we can not be morally responsible for them.

IV. Frankfurt and the Principle of Alternative Possibilities

Harry Frankfurt argues against incompatibilism as represented in the consequence argument.
He argues that free will, and specifically moral responsibility, is compatible with Laplacian determinism.

Frankfurt begins by noting that we are inclined to endorse the following principle of alternate possibilities (PAP):

> A person's act is free if and only if that person could have done otherwise (829).

On PAP, if determinism is true and incompatible with free will, no one ever could have done otherwise.
No one ever acts freely.
And, thus, no one can be morally responsible in a deterministic universe.
Frankfurt argues that one can be morally responsible even if one could not have done otherwise.
One can act freely even if one could not have done otherwise.

Frankfurt presents the example of $Jones_4$, which seems to provide a counterexample to PAP.

> Suppose someone — Black, let us say — wants $Jones_4$ to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until $Jones_4$ is about to make up his mind what to do, and does nothing unless it is clear to him (Black is an excellent judge of such things) that $Jones_4$ is going to decide to do something other than what he wants him to do. If it does become clear that $Jones_4$ is going to decide to do something else, Black takes effective steps to ensure that $Jones_4$ decides to do, and that he does do, what he wants him to do... Now suppose that Black never has to show his hand because $Jones_4$, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, $Jones_4$ will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. It would be quite unreasonable to excuse $Jones_4$ for his action...on the basis of the fact that he could not have done otherwise. This fact played no role at all in leading him to act as he did... Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it (835-6).

So, $Jones_4$ could not have done otherwise, since Black was prepared to force him to act, but $Jones_4$ still bears moral responsibility.
Note that Black, in this example, is a stand-in for the laws of physics.
He is what ensures that $Jones_4$ could not do otherwise.
While Black was not impelling $Jones_4$ to act, he was ensuring that $Jones_4$ could not have done otherwise.
Yet, $Jones_4$ was responsible for his action.
Thus, PAP seems false.

Frankfurt has shown PAP to be false without impugning the more plausible claim that moral responsibility is excluded by coercion.
If we are truly coerced, we are not morally culpable for our actions.
But, there are cases, like that of $Jones_4$, in which we can not do otherwise, and yet we are morally responsible.

Frankfurt suggests a revision of PAP:

> A person is not morally responsible for what he has done if he did it *only* because he could not have done otherwise (838, italics added).

This new principle is consistent with the claim that we are not responsible when coerced.
Further, it leaves moral responsibility compatible with determinism.
We might act in a particular way and have no other options, as the determinist says.
In such a case, we act *only* because we could not have done otherwise.
Still, we are not necessarily off the hook, morally, if we wanted to act as we did.

> The following may all be true: there were circumstances that made it impossible for a person to avoid doing something; these circumstances actually played a role in bringing it about that he did it, so that it is correct to say that he did it because he could not have done otherwise; the person really wanted to do what he did; he did it because it was what he really wanted to do, so that it is not correct to say that he did what he did only because he could not have done otherwise. Under these conditions, the person may well be morally responsible for what he has done (839).

One question for Frankfurt, it seems, is whether the case of $Jones_4$ can be extended to a general defense of compatibilism.
Is it merely an obscure counter-example?
Or, does it demonstrate a general fault with the incompatibilist's position?

Another question, and this one is bugging me: Is Frankfurt's compatibilism any improvement on the old-fashioned Humean version?
That is, does it take seriously the metaphysical problem that our wills also seem determined?

V. Identification, and the Three Questions

In Frankfurt's compatibilism, coercion excuses moral responsibility.
If we act only on our will, though, in the absence of coercion, we may be taken to be responsible.
Frankfurt's case of $Jones_4$ showed that even if we can not act otherwise, even if the universe is determined, as long as our reasons for acting were derived from our will, rather than from external coercion, we can be (or are) held morally responsible.
In later work, Frankfurt explores the notion of identification.
We identify with a behavior when we embrace it, or do it in accordance with our will.
Thus, we can say that Frankfort ascribes responsibility for an action as long as we identify with that action.

Still, Frankfurt originally agreed that when we are coerced, as in the case of $Jones_2$, we are not responsible.
Note that if we are committed to a behavior, or type of behavior, we can identify with it whether or not we are coerced into doing it.
Woolfolk, Doris and Darley look at attributions of responsibility in cases of coercion and find that the case is not as simple as Frankfurt originally claimed.
Frankfurt assumed that coercion excuses moral responsibility.
But it looks as if ascriptions of moral responsibility are higher if we identify with an action.
We may be held responsible, even in cases of coerced action.

> Persons may be held responsible even when they operated in the grip of forces compelling that action (64).

We can call the adjusted view mature compatibilism.
Remember, that compatibilism is a way of rectifying the apparently conflicting claims of strict causal (or hard) determinism with the recognition that there are certain actions for which we ascribe responsibility to an agent, and others for which we do not.
We have intuitions that certain agents are responsible and others are not.
Woolfolk et al. seek experimental confirmation of the intuitions that ground mature compatibilism

Further, Woolfolk et al. are interested in psychological research which relies on a causal discounting principle.
The causal discounting principle, which is consistent with Frankfurt's original compatibilism, claims that people's attributions of moral responsibility are discounted when an agent is perceived to be constrained.
That is, psychologists assume that coercion excuses responsibility.
The causal discounting principle is applied outside of psychology since questions of moral responsibility

are required when determining legal punishments.
If the mature compatibilist is correct, then the causal discounting principle will not be defensible.
For, one has to consider not just whether the agent is constrained, but also whether s/he identifies with her/his actions.

Lastly for background, there is an assumption, among some philosophers, that the folk are naturally incompatibilists.
Remember, there are two kinds of incompatibilists.
Libertarians argue that we are actually free; they deny determinism.
Hard determinists accept determinism and infer from it that people are never morally responsible.
If people ascribe responsibility in cases of maximal constraint, there would be experimental evidence against hard determinism that could not be explained by libertarianism.
Thus, one might demonstrate that people are naturally compatibilists.

Woolfolk et al. thus seek data concerning three questions:

1. Are folk theories compatibilist or not?
2. If they are compatibilist, are they mature compatibilists, ascribing responsibility even in cases of constraint? Or, do they, as Frankfurt originally did, excuse coerced action?
3. Is the causal discounting principle consistent with folk intuitions?

VI. The Woolfolk Experiments

Much of the Woolfolk et al. paper consists of descriptions of their carefully crafted research designed to help answer the three questions above.
They construct stories with high and moderate constraint, and with high and low identification.
All stories involve a cuckold and a hijacking.
In high identification, Bill wants to kill Frank because Frank has been having an affair with Bill's wife.
In low identification, Bill finds out about the affair, but comes to grips with it, and is at peace with Frank.
In moderate constraint, Bill is asked by hijackers to kill Frank, but he can (possibly) avoid it.
In high constraint, Bill is threatened by the hijackers and it seems that the only way to avoid killing Frank is to die himself.

To respond to criticisms that defying the hijackers is an option, even in cases of high constraint (i.e. that the situation of high constraint is not really one in which one is coerced), Woolfolk et al. devise a further absolute constraint, under which Bill is given a psychoactive compliance drug which renders him physically unable to avoid complying with the hijackers' requests.
Subjects given the absolute constraint scenario were additionally told to suspend their disbelief about the plausibility of the compliance drug.
It might be interesting to pursue a question about the applicability of results which require a suspension of disbelief; it could be the case that the folk don't really have a notion of absolute constraint applicable to cases such as these.
In other words, if people adopt compatibilism as an alternative to hard determinism when they are told that libertarianism is not an option, the result would not confirm that those people are compatibilists; it would only show that they prefer compatibilism to hard determinism.
People might prefer libertarianism to compatibilism.

Lastly, Woolfolk et al. performed a separate experiment (Experiment 3, pp 73-75) to test whether subjects generally attributed greater responsibility in situations of lesser constraint.
They did.
The point of this last experiment was just to verify that any effects of identification are working against a backdrop of a general (folk) theory which negatively correlates constraint with responsibility.

To judge whether subjects attributed responsibility to Bill, when he kills Frank, Woolfolk et al. asked them directly whether they thought Bill was responsible.
In addition, they probed the subjects' thoughts about responsibility with further, related questions (p 67, Table 4.1).
Some of these factors were combined into a further measure, called Bill's blameworthiness, and into an opposing measure, called hijacker responsibility.
As expected, they found that Bill was deemed significantly less responsible in cases of higher constraint.
More interestingly, Woolfolk et al. found that subjects deemed Bill more responsible in high identification cases than in low identification cases.

> We found participants' attributions of responsibility for an action to be influenced by the actor's attitude toward that action, even when the action was causally constrained to such a degree that there were no other behavioral options. The degree to which actors "identified" with an action was strongly associated with responsibility for the action being assigned to them (75-6).

VII. Descriptive Results and Normative Questions

The three questions which Woolfolk et al. raised with regards to their research were:

1. Are folk theories compatibilist or not?
2. If they are compatibilist, are they mature compatibilists, ascribing responsibility even in cases of constraint? Or, do they, as Frankfurt originally did, excuse coerced action?
3. Is the causal discounting principle consistent with folk intuitions?

The results show clearly that folk intuitions are more complex than the causal discounting principle may indicate.
Identification, in addition to constraint, affects our attributions of responsibility.

> Our data suggest that the attribution of moral responsibility takes into account noncausal elements, such as identification, in addition to causal factors... Our data support the view that information about outcomes that an actor desires can moderate or override the attributional effects of the actor's perceived control over events (76).

Woolfolk et al. make further, reasonable suggestions for psychological research into folk theories of responsibility.

More interesting, for us, are responses to the first two questions, concerning the philosophical questions about free will and compatibilism.
Woolfolk et al. argue that their data show that, in some contexts, people are not incompatibilists.
We attribute some level of responsibility to Bill, even in the situation of absolute constraint, in situations of high identification.

Woolfolk et al. take their work as having shown that folk theories reject the principle of alternate possibilities.
But the results are even stronger.
The data support a mature compatibilism.
People may find others responsible even when they are stampeded by a threat.

But, Woolfolk et al. also deny that folk theories are strictly compatibilist.
They argue that people do not have a simple theory of responsibility attribution.
Instead, they allege, people are contextualists: the amount of responsibility that people attribute depends on a constellation of factors particular to the case.

> Differing considerations are salient to moral responsibility attribution in different contexts, and...patterns of responsibility attribution may also vary culturally and developmentally. We might suspect that along the broad spectrum of moral cognition related to responsibility ascription, which ranges from the determination of criminal liability, to the assignment of credit for scientific discovery, to deciding which sibling should have to clean up the spilled milk, complex considerations often come into play (77).

It is clear that research into people's actual attributions of responsibility introduce complexity into our analyses of responsibility.
We have returned, thus, to the tension between a normative theory and a descriptive theory.
In constructing theories, we aim for simplicity.
But, the simplest theories will have to smooth out the rough edges found in studies such as these.
When we smooth over the rough edges, do we irresponsibly ignore data in favor of our intuitions?
Or, do we take a normative perspective in which the actual attributions of responsibility are evaluated according to a prior standard?

Here is another way to put the question: does folk-psychological rejection of the causal discounting principle indicate that we should ascribe more responsibility in cases of higher identification, even in cases of maximal constraint?
Perhaps the folk just get it wrong.