

[[Connect To TPM Online](#)] [[Instructions](#)] [[Return to Main](#)] [[Articles](#)] [[News](#)] [[Features](#)] [[Essays](#)]
 [[Conference Reports](#)] [[Open Debate](#)] [[Primers](#)] [[Forums](#)] [[Reviews](#)] [[Regulars](#)]
 [[Philosophy on the Web](#)] [[Keywords](#)] [[Search The CD-ROM](#)]

Determinism and Punishment

An Interview with Ted Honderich

Stangroom: What is determinism? And what is its philosophical significance?

Honderich: Determinism as I understand it is the doctrine that each of our mental or conscious events or episodes, including every decision, choice and action, is the effect of a certain kind of causal sequence. The sequence goes back a long way in time before the decision, choice or action. Also, the sequence is one of standard causation. Each event in it is a real effect – a necessary event, so to speak. Such an effect is not an event merely made probable by antecedents. It is something that had to happen given the antecedents as they were. Determinism, so defined, isn't itself the doctrine that we are not free – that question is not touched on in this definition.



Stangroom: Why does it not assert that we are not free?

Honderich: Well it's been thought to be an open question whether or not freedom is inconsistent with such a doctrine of determinism. There are philosophers - at one time they seemed to be a majority in philosophy in the English language – who have taken the view that determinism and freedom are logically consistent, that determinism can be true and yet we can still be perfectly and entirely free. So it's a good idea – anyway convenient - to leave anything about freedom out of a definition of determinism.

Stangroom: This is the Compatibilist position?

Honderich: It is indeed.

Stangroom: In this respect, the distinction between *origination* and *voluntariness* is significant. What is this distinction?

Honderich: Compatibilist philosophers, as already remarked, have defined freedom in such a way that it is logically compatible with determinism. This freedom, what I call freedom as voluntariness, essentially amounts to this: a free action is one that flows from the desires, personality and character of the agent, rather than somehow being against those things. The agent is not in jail, not subject to a man with a gun, not subject to a compulsion internal to him. He is acting in such a way that his actions in a clear sense flow from himself. A free action by this definition is indeed logically consistent with determinism. Determinism doesn't say that there are no actions that flow from the agent. It just says that there is *some* causal background that fixes the outcome. A free action on the Compatibilist account is just is one that has a certain kind of causal background -- internal to the agent, so to speak, rather than external.

As for origination, it comes from the opposing tradition in philosophy that maintains that a free action is inadequately defined by the Compatibilists. Kant asserts that Hume, in taking up the Compatibilist definition of a free action, is merely playing around with words, merely on the surface of things. For Kant and the Incompatibilists, a free action is a voluntary one but also much more than that. It is one that has a certain genesis, a certain inception, rather difficult to define. In one sense, we know what Incompatibilists take origination to be. It's the agent coming to a decision, choice or action in such a way that determinism is not true of this, and yet the decision, choice or action remains within the control or devising of the agent. Above all, origination is a beginning of a decision or choice that makes the agent responsible for it – morally responsible for it in a strong sense. Free actions, if they are both voluntary and originated, are certainly inconsistent with determinism. If determinism is true, there aren't any of these free actions.

Stangroom: Are theories of originated action coherent? Is it possible to specify what an originated action might be? It seems quite hard to imagine such a thing.

Honderich: Many philosophers have said there is the greatest difficulty about arriving at a clear conception of origination. Part of the difficulty is that an originated decision, if there are any, is one that could have been different at the very moment it is made. If I decide to shoot Thatcher, I can at that moment decide to act differently. What this means is that the past could have been exactly as it was up until that moment and I can nonetheless decide differently. That story contains within it a pretty alarming proposition, or it seems to --namely that *there is no explanation of the decision I in fact make*. You have to look for such a thing look in the present and past circumstances. But in both of these places, as we've been hearing, everything could have been the same and I could have decided differently. So in one clear sense, it appears, there is no possibility of any explanation whatever of this decision that was made.

But that isn't to say that *no* sense can be made of origination. After all, I can define it as the giving rise to a decision in such a way that the decision is not determined, and yet is within the control of the agent, and moreover in such a way that he is really responsible for it. I haven't said nothing when I've said that. I haven't said something incoherent. I haven't explained how there can be such decisions, but I have said something that appears to make sense. Indeed, what we have here is an idea or conception in common usage. People sometimes think after someone has behaved very badly, maybe viciously, that the person could at the moment have stopped doing the thing, given things just as they were. So the idea of origination, even if it does contain a mystery, does exist. Sense can be given to it, and it seems to be entrenched in ordinary culture – anyway Western culture as we know it.

Stangroom: So in this conception, how is it imagined that the agent escapes the network of causality which seems to exist with respect to all other material – and indeed, non-material – phenomena.

Honderich: I have to admit again being unable to give a good answer to that. But that doesn't commit me to thinking there doesn't exist any conception of origination. There *can* be conceptions that are partly mysterious. There are lots of them. Of course, I don't think there really are things of which the conception of origination is true. What is true is determinism. It seems you agree with that, but we are in a minority. Most people are inclined to think that determinism is false. They talk of Free Will and have in mind something like the origination we've been talking about – anyway a kind of image along those lines. Also, there are a lot of more informed characters about who

have heard of Quantum Theory, the physics of this century. They think it refutes determinism.

Stangroom: I've never really understood that argument. I've always thought it something of a red herring. Is that your view?

Honderich: In my view it is pretty hard to make into a real argument. My own resistance to this idea that Quantum Theory falsifies determinism has got at least two parts.

Firstly, if there really is indeterminism – uncaused events, events that aren't effects – then they are of course at a micro-level, well below the level, for example, of brain events that go with choices and decisions. More important, they don't translate upwards to the macro-level. That is our experience. We don't see miraculous little events, chance events, like spoons levitating. We ought to have this evidence if the miraculous micro-events come up to the top. So a first resistance to the Quantum Theory stuff is that if determinism is true, it's irrelevant.

My second resistance is to their actually being any of the events in question, down at the micro-level. All the popular books about Quantum Theory, some of them by distinguished physicists, say one thing. It is that you can't carry over old assumptions from classical physics into contemporary and recent physics. One of the things that you can't carry over is a conception of the nature of the things that before Quantum Theory used to be said to be caused or determined. For example, it is said that if the term "particle" is used in an interpretation of Quantum Theory, you are not to suppose a particle is a small bit of matter in the Newtonian way. It is very uncertain in the end, and indeed this is admitted by most exponents of Quantum Theory, *what* the things are that are said not to be effects. Sometimes they are taken to be probabilities or possibilities or indeed propositions.

The essential point here is that it looks like the things that we are told are not effects are things that the determinist never said were effects. No sensible determinist has said that numbers – say the numbers 4 and 5 - are effects, or that propositions are effects. These are thought to be abstract objects and no one has supposed that the determinist is committed to saying that these are effects. No one supposes that a determinist is committed to saying that a space-time point - for example, the space-time point at the end of my finger - is an effect, and the determinist doesn't say that it is. Determinism, plainly, is only about *events*, or a certain class of *events*. In short, to repeat, it's very possible that the things that are asserted in Quantum Mechanics not to be effects are in fact not events at all, and are therefore not relevant to determinism.

Stangroom: You believe that both Compatibilism and Incompatibilism are inadequate. Why?

Honderich: I actually believe I've proved both are false. Is any philosopher allowed one moment of pomposity? I'll use mine up here. I can't say Compatibilism has gone away. I heard a lecture on it the other month at the Royal Institute of Philosophy that sent all persons present sound asleep. But it should have gone away, like Incompatibilism.

What my stuff comes to can be put in terms of life-hopes, certain attitudes to the future. We all have them. These are large hopes about the working out of our lives.

They have to do with our future actions and what will flow from those actions. What is most important, however, is that these particular attitudes come in two kinds. You can discover the two kinds in yourself.

I can feel about the future in a way that makes it bright. The essential point is that I can have an attitude to it as something in which I will get what I want, where I'll be doing what I want. I'll end up with the right person, or with money, or just healthy, or whatever. I won't be alone or in jail or bed-ridden and so on. Things will turn out in accord with my personality and nature. If I'm in this mode of feeling, furthermore, I can feel that determinism can turn out to be true, and it won't matter much. All of us have this kind of hope, or at the very least can get into it.

On the other hand, almost all of us have or can get into a very different sort of hope. It's to the effect that we're going to be able to rise up over our pasts, rise over our characters, rise over our weaknesses, and defeat the things which have kept us back – anyway to some extent. Our futures aren't written down waiting to be read, fixed already. This is a hope, further, that is wrecked if we think of determinism as true.

That we have or can have both these attitudes shows that we have both the conception of free actions as just voluntary and the conception of free actions as both voluntary and originated. The first conception, plainly, is in the first sort of hope, and the second conception in the second sort. Both these ideas are within us. If that is true, then both Compatibilism and Incompatibilism are false. They are both false because they agree in one thing – that each of us has one single conception of a free action. The Compatibilists say that it is voluntariness and the Incompatibilists that it is voluntariness plus origination. They're both up the spout.

By the way this isn't just *asserting* that 'free' is ambiguous between the Compatibilist and the Incompatibilist definitions. We can show that we have two different sorts of attitudes, different in their contained feelings. In fact they are connected to some extent with different behaviour. The two sorts of attitudes encapsulate the two ideas. One is of actions as only voluntary, and one is of actions as both voluntary and originated. The existence of the attitudes makes for something like a proof of a behavioural kind that people have an idea.

If Compatibilism and Incompatibilism are both false in this way, the real problem of determinism of course isn't what our single shared idea of freedom is – we've got two. The real problem of determinism is living with and somehow emerging from the situation where we've got two conceptions of freedom and they enter into important attitudes that we have – our life-hopes and a good deal more.

Stangroom: What about an Incompatibilist who is also a determinist? Such a person allows that voluntariness is compatible with determinism, but maintains it has nothing to do with freedom or anyway isn't all of freedom. So the short story is that we're not free and responsible.

Honderich: Well, a determinist can say that determinism is compatible with voluntariness, and voluntariness is freedom, and so everything is okay. Or, as you imagine, a determinist can say that origination is needed for freedom, and so since determinism is true, there isn't any freedom. My line is that each of us, if we come to believe determinism, are in something like both of those positions. As for the second one, we are now all inclined to feel that if someone does us a tremendous benefit, in adverse circumstances, they could have done otherwise in a real sense. That is why

we're grateful. If we get converted to determinism, we'll have a problem.

Stangroom: We don't feel grateful to machines. Determinism if it's true turns us into biological machines, doesn't it? What room is there for any kind of gratitude at all there?

Honderich: I agree that we don't have certain desires in connection with machines. When they benefit us, we don't have certain desires somehow to do well by them in return, if only by saying thanks. We don't have the counterparts of the retributive desires we may have if a person wounds us. But it seems to me that there is another kind of feeling we can have in connection with a machine that benefits us, a good feeling. There is a relation of that feeling to a larger thing – something we could still have to people if we took determinism to be true.

Stangroom: You talk about *dismay*, *intransigence* and *affirmation*, as responses to determinism in this connection. What are these things?

Honderich: Dismay is a response to determinism that may have to do with life-hopes, claims or feelings of knowledge, personal feelings, moral approval and disapproval and so on. Dismay is the response that if determinism is true, these things are wrecked. My life-hopes must collapse, and so on. I can't be confident in what I used to call my knowledge. I can't engage in gratitude or resentment. I can't hold people responsible.

Intransigence is the response that if determinism is true I can still soldier on -- with my life-hopes, personal feelings and so on.

These two responses, which people demonstrably have, also entail that they the conception of freedom as voluntariness together with origination, which is inconsistent with determinism, and also the conception of freedom as just voluntariness.

Being inclined to both these responses is no happy thing. You're in a kind of conflict situation for a start.

What is needed is to make the response of *affirmation*, which you might think boils down to getting rid of desires that cannot be satisfied if determinism is true and being as fulfilled as possible in the fact that other desires still can still be satisfied. Something better can be said along those lines. Affirmation can be the response that life can be great and fulfilling. As for the giving up on the other desires, the best way to succeed in it is to come to believe in determinism.

Stangroom: Moving on to punishment, what is it?

Honderich: Well, an infliction of a penalty on an offender by an authority. An imposition of something undesired on someone who has broken a rule, and an imposition by someone or something who is empowered to do it by another rule. What is left out of that definition, something that traditionally is put in, is "for an offence". If you put that in, then from the beginning you beg the question as to what punishment is for -- what the rationale or justification of punishment is. Therefore it's a good idea to leave that out.

Stangroom: What is a "justification", philosophically speaking?

Honderich: I take it that justification is a good reason – maybe a sufficient reason – for thinking that something is right. There seem to be three large questions in morality and moral philosophy. One is the question of right actions, what actions or practices are right -- particular actions or classes of actions. The second question is what persons are okay or estimable or human -- the question of good agents. It's about people over stretches of time, or with respect to their whole lives. The third is the question of moral responsibility – appraising and feeling about persons with respect to particular actions. The problem of the justification of punishment is a problem in the first category and asks you to give a sufficient reason for the rightness of the practice of punishment by the state.

Stangroom: Are there any criteria for an adequate justification of punishment?

Honderich: Well for a start anything offered as its justification has to be conceptually adequate -- it has to be decently clear.

What is offered in the Retribution Theory is to the effect that punishment is justified because it is deserved. What does it mean? There have been many attempts to make sense of the claim that punishments are deserved and most of them are failures. Judges persist in saying or seeming to imply that there is some kind of factual equivalence between the distress of a certain penalty, say ten years in prison, and the culpability in the offence, where the culpability is the amount of harm caused and the degree of responsibility. But of course there is no serious possibility of quantifying the distress of the penalty and the amount of harm and the degree of responsibility. And very certainly there are no common-units for the quantification of these three things. So it makes no sense to talk about equivalence between penalty and culpability in this way.

What I say in the end is that the Retribution Theory does make sense in another way. To speak of a penalty as equivalent to an offence is to say that a certain penalty will do no less than, and do no more than, satisfy the grievance raised up by the offence. The grievance is a desire, a desire for the distress of another person. Not as a means to something else, but just for itself. It is the low desire that someone else should be distressed.

Stangroom: Whatever else is to be said of that, doesn't it raise an "Is/Ought" problem?

Honderich: May I first say a couple of other things? One is that it's obvious that the Retribution Theory doesn't work. I don't think for a moment that it does. If all you can say for putting a man in jail for ten years is that this satisfies a grievance-desire on the part of the victim or the family of his victim, that doesn't seem to be enough. A second thing has to do with the idea that the existence of such a low desire is no reason at all for doing anything. Some philosophers say that sort of thing. I'm inclined to think, on the contrary, that the very existence of any desire is a reason for satisfying it. The reason may well be entirely outweighed by something else, of course, as in the case of putting the man in jail for ten years.

Stangroom: But why ought we to do what satisfies desires? How can the second thing entail the first? How do you get an 'ought' from an 'is'?

Honderich: I agree there's no logical entailment. But it's also true, maybe, that there can't be any greater reason for an action than that it does satisfy a desire – say the

desire to be relieved of pain and suffering or victimization or injustice or whatever.

Stangroom: So you would be prepared to say that, for example, in Nazi Germany, the existence of the desire for the distress of Jewish people was a reason for its satisfaction?

Honderich: Yes, that is exactly the position that I'm committed to -- and, needless to say, I rush to add that it was a desire that couldn't have begun to justify what happened, since what was done was something whose frustration of other desires was overwhelming, appallingly so.

As for 'is' and 'ought', you say I'm arguing from the first to the second. Indeed I am -- and I suppose that puts me in the company of virtually all persons who engage in moral judgements. That an "ought" cannot be deduced from an "is", which is plainly true, certainly doesn't give us the conclusion that factual reasons are impossible for moral conclusions. Suppose we're now faced by a suffering child. I say we've got to do something, and you say we've got to do something. If we suddenly have the idea that the fact that it is suffering doesn't entail that we've got to do something, we're not going to just go for a walk instead. We've got an excellent reason for doing something, even though the excellent reason doesn't entail the evaluative conclusion.

Stangroom: But you've got no way of *demonstrating* to someone who claims that the suffering child should continue suffering that their position is incorrect.

Honderich: Well it seems to me that we can reflect pretty conclusively on that subject. Here's an old example of mine, the two-button box. It's like a telephone kiosk. People who go into the box face two buttons. One is a pain button, one is a pleasure button. Or say one is the contentment-and-fulfilment button and the other is the distress-and-deprivation button. The important thing about the box is that the persons in the box are also in a state of epistemological deprivation. They don't know anything about the effects of pushing the buttons, except that pushing the pleasure-button will cause pleasure to someone and pushing the pain-button will cause pain. They don't know anything else. They haven't any idea, for example, whether pressing the pain button, although this causes some pain, is the way to prevent a lot more pain in the future. And they don't know anything about the pleasure button except that pushing it will cause pleasure. They don't know anything about the person who will get either the pain or the pleasure, and so on.

Now consider someone who in that situation understands all this and knowingly presses the pain button. What is to be said of him?

Stangroom: Well, we'd say that he was mad.

Honderich: Yes, I agree. Or you might say he wasn't human.

In this situation we don't seem to be ready to say that morality is a matter of taste, that it is just a matter of attitudes, that it is all in the eye of the beholder, or any of that kind of thing. In this clarified situation, for someone actually to choose pain and degradation over pleasure and fulfilment would count, as you were the first to say, as mad. Where to go from there I'm not quite sure. But it does give one a kind of handle on a stropy undergraduate who says it's all a matter of attitude -- or indeed someone who says, truly, that you can't *demonstrate* moral conclusions. You *can* think they have a kind of foundation in our sanity and humanity.