AS Philosophy: Free Will and Determinism

PHILOSOPHERS ON STRIKE

SUPPORT EPISODEMOLIGISTS LOCAL 191
UNITED FEDERATION OF PHILOSOPHERS

NO MORE SEARCH FOR TRUTH UNTIL
OUR DEMANDS ARE MET.

WE ARE RIGHT
AND I CAN LOGICALLY PROVE IT.
Introduction

The “Free Will” topic deals with the question of whether our concepts of human freedom and moral responsibility are “compatible” with the way the world is. In other words it asks whether, once we fully understand the way the universe works, and our place in it, we are able to carry on thinking of ourselves as beings who make free choices, and who are responsible for our actions.

One aspect of “the way the world is” is particularly important for this topic: this is the thesis of determinism, which is the idea that everything that is going to happen in the universe is fixed in advance by what has already happened plus the laws which govern how the universe works. Determinism is a controversial thesis – scientists continue to debate whether the behaviour of every particle in the universe is controlled by “deterministic laws” (laws which fix how that particle must behave in every possible circumstance); and philosophers debate whether the human mind should be thought of as governed by the same kind of rules that govern the behaviour of physical objects like electrons, quarks, and neutrons.

The structure of the debate here is simpler than that of the “Reason and Experience” topic: it divides into just two main questions:

- What is the relationship between determinism and human freedom?
- What is the relationship between determinism and moral responsibility?

Nevertheless, the terminology can be confusing at times – trying to memorize the meanings of key terms as you go along will help you stay on top of the material.
1 – Some Basic Distinctions

Aims: to understand the following distinctions:

- between actions and bodily movement
- between reasons and causes

To understand what is meant by the terms “mind” and “body”.

We inhabit a physical world, and we might think of ourselves (or at least our bodies) as a certain kind of physical object – one composed of flesh, blood, bones, brain cells and so on. However, we have some concepts which seem to apply only to us (and other people), and not to physical objects such as tables, chairs, rocks, and so on. If we can understand these concepts, we might hope to get a better idea of why we describe people as acting **freely**, and making **free choices**, while a stone or a chair never counts as choosing to do anything.

In particular, people (and not chairs) can do things – they can “act” or “perform actions” (n.b. philosophers describe anyone who acts in this sense as an “agent”). Moreover, when people act, they do so for reasons – we are not simply caused to act by our circumstances, but weigh up different courses of action rationally. Finally, we think of people as having **minds**: although some philosophers argue that “mind” and “brain” are two different words for what is really the same one thing, it seems that we describe minds in radically different terms from brains, and this has led many people to claim that mind and body are really **distinct**; or at least that we should not mix up the two ways of talking.

You should prepare yourself to **explain and illustrate** the distinctions between reasons and causes, and between action and bodily movement; you should also get used to making a distinction between vocabulary appropriate to **minds**, and vocabulary that applies directly to physical **bodies** or **brains**.

**Exercise:** without reading further, try to think of, and write down, three differences for each of the two distinctions – actions and movements, and reasons and causes.

**The distinction between reasons and causes:**

- Causes explain events by connecting them to the laws of nature: the event of the rock hitting the antique vase caused the event of the vase shattering, because the laws of nature determine that any object with the same physical properties as the vase will shatter when hit by an object with the same physical properties as the rock. On the other hand, reasons explain actions by showing how those actions follow from the beliefs, desires, and values of the agent, and from the standards (or “norms”) and values that obtain in the agents’ society.
- This point is sometimes made by saying that thinking of someone as acting for a reason is part of taking what Daniel Dennett termed the “intentional stance” – i.e. it is part of the way of thinking which predicts human behaviour in terms of reasons, desires, and rationality, rather than in terms of the “mechanical stance” of predicting events by examining causes and laws. Here’s how Dennett explains it:

  “Here’s how it [the intentional stance] works: first you treat the object whose behaviour is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set
of beliefs and desires will in many – but not all – instances yield a decision about what the agent ought to do; that is what you predict the agent will do.”


- Reasons (and not causes) are **goal-orientated**: to act for a reason is to have some specific goal or purpose in sight (even if this purpose is to prevent something from happening). An event which is merely caused by another event doesn’t have to have a purpose.

- We don’t generally have to be aware of, or able to find out, the causes of an event, while we are typically aware of the reasons for our actions – or at least we can generally supply our reason(s) for acting as we did if challenged. (n.b. some would argue that our real reasons for acting can be very different from the ones we provide when asked; they would say that our real reasons only become apparent to us through therapy and/or psychoanalysis. Nevertheless, they may still accept that the “real reason” is always in principle discoverable.)

- Doing things for reasons is a necessary condition for counting as an agent (someone who performs actions) in the first place; something that does things simply because it is caused to do so (like a machine) doesn’t strictly speaking perform actions or “act” at all.

- One can have both good and bad reasons for actions, while causes are immune to criticism: it is possible to criticize an agent for acting on a reason that she ought not have acted on, while it makes no sense to criticize an event on the basis of what caused it to happen.

You should be aware that some philosophers (e.g. Donald Davidson) have argued that reasons are causes: that my reason for performing an action is what causes me to perform that action. If we are convinced that the cause of my action is an event in my brain (e.g. a neuron firing), then we might be led to say (as Davidson does) that when we describe the reason for my action we are really just giving another name for the event in my brain that caused my action. Nevertheless, the range of differences between reasons and causes given above makes it very controversial to claim that reasons and causes are identical.

**The distinction between action and bodily movement:**

Notice that “bodily movement” here means a physical movement of the body, described in purely physical terms: for example a change in the position of my left arm caused by neurones firing and muscles contracting. Thus an important distinguishing feature of bodily movements is that they can only be described in the language of science; once we use vocabulary that implies a social or cultural context we are describing an action, not merely a bodily movement. Other important differences are as follows:

- One bodily movement can count as several different actions: for example, the act of raising my arm might simultaneously be the act of volunteering to answer a question, and the act of obscuring the view of the person behind me.

- The reason for this is that we distinguish and explain actions within a social or cultural context: an act such as bidding at an auction (again, by raising your hand) can only take place within an appropriate social context, and can only be recognized and understood by someone who understands that social context.

- Another aspect of this is that we can describe actions (and not bodily movements) as “meaningful” or “significant”. For example, the act of winking at someone has a meaning – although of course this meaning depends on the social context.

- Some actions are passive: think of sitting down, or waiting for a friend.

- Describing someone as performing actions involves us in taking the “intentional stance” towards them: i.e. treating them as rational beings motivated by reasons and desires. If we ask someone
why they acted as they did, we expect a statement of their reasons; while an explanation of a bodily movement proceeds in terms of causes and laws of nature. Similarly, we think of actions as carried out with some purpose in mind, while physical bodily movements need not have any such purpose.

- We can be held accountable for actions, but not for mere bodily movements – for example, I can be held responsible for causing a car crash by driving too fast (an action), but am not morally (or legally) responsible for a car crash which is caused by my having a sneezing fit (a succession of bodily movements which do not count as an action).

- It makes sense to distinguish between “voluntary” and “involuntary” actions: remember the auction? I might have voluntarily performed the action of waving at a friend – and at the same time involuntarily performed the action of bidding a million dollars for an art work. The terminology of “voluntary” and “involuntary” does not seem appropriate for mere bodily movements (unless we are also thinking of them as actions).

**Mind and Body**

The precise nature of the relationship between minds and bodies is hugely contentious (in fact, it forms a large part of the A2 course); however, you need to understand a little bit about the apparent differences between the two. Notice that “body” in this sense means a physical object, described in the language of science – while when we describe our “minds” we use the language of beliefs, desires, feelings, and reasons. Strictly speaking, then, we should try to avoiding talking about our brain (a physical object) using the language appropriate to our minds – at least until we have an adequate philosophical theory of the relationship between the two.

**Exercise:** which of the following statements would you count as inappropriate uses of language?

1. “My brain is hungry”
2. “My mind is composed of billions of neurons”
3. “My brain hurts!”
4. “I am in love with her brain”
5. “I want my brain to be happy”
6. “I want my brain to be well-nourished”
7. “I want my mind to be well-nourished”
8. “I know everything that is in my mind”
9. “This pill will make your brain cleverer!”
10. “I’m not responsible for killing him – but my brain is!”
2 What is determinism?
Aims: to understand what is meant by “determinism”. To distinguish determinism from fatalism, predestination and predictability. To understand how subjective probability (“chance”) may be compatible with determinism.

Defining Determinism
Determinism is a thesis about the way the (physical) world is: it is the claim that the world is set up such that “a determinate set of conditions can only produce one possible outcome given fixed laws of nature.” To understand that better, think about the kind of world view that the determinist denies: call this the “branching futures” model. According to this, although the past and present are fixed (they cannot change, as they have already come into existence), the future is not yet fixed: there are many possible ways things could turn out in the next hour, and even more possible ways the world could develop in the next century. Notice that the “branching futures” model doesn’t describe merely our state of knowledge of the future – everyone agrees that we don’t know exactly what is going to happen next. Instead it is a thesis about the world – that the way the future is has not yet been fixed, and there are many different futures, each of which is genuinely possible at this moment.

The determinist rejects that model of how the world is; instead a determinist will say that there is only one possible future, given the way things are now. The reason for this is that the determinist believes the world is governed by fixed, exceptionless laws (the “laws of nature” or “laws of science”) which determine or fix what will happen next, based on the current conditions in the universe. We could put it like this:

The current state of the universe
+ The laws of nature
  determine
The state of the universe at every point in the future.
You should notice an important consequence of this deterministic way of thinking about the universe: because, according to the determinist, the total state of the universe at any time determines the state of the universe at any later time, everything that has happened, is happening, or is going to happen, was already fixed by the conditions at the very beginning of the universe plus the laws of nature. So 13.75 billion years ago, at the very start of the universe, the initial conditions of the universe fixed what the weather would be like in 2012 (plus what exam grades you are going to get this summer).

That brings up an important question: if determinism is true, does it cover our own actions as well as the behaviour of physical objects without minds? The answer is that it seems like it has to, given what we know about how our actions depend on physical activity in the brain. Everyone agrees that you can’t go for a walk, raise a hand, or write an answer to an exam question, without millions of neurons firing in your brain to control the muscle groups involved; and since those neurons are physical objects (a certain kind of cell), their behaviour ought to be governed by physical laws just like anything else that is made of electrons, protons and neutrons. If the determinist is right, and those physical laws or “laws of nature”, fix everything that is going to happen next on the basis of the current state of the universe, then those laws also fix which neurons are going to fire in your brain over the course of your life, and so fix everything you’re going to do.

Determinism has been endorsed by many (but not all) famous scientists. Here’s how Albert Einstein put the point:

“Everything is determined, the beginning as well as the end, by forces over which we have no control. It is determined for the insect as well as for the star. Human beings, vegetables, or cosmic dust, we all dance to a mysterious tune, intoned in the distance by an invisible player.”

Einstein’s “invisible player” here is the set of laws which govern the universe – and the “mysterious tune” is the complicated behaviour of the micro-physical particles which control the behaviour of the objects (including us) that they make up.

**Determinism distinguished from fatalism:**

It is important not to confuse determinism with other, less plausible, views which also claim that our future is fixed. The determinist claims to know that the future is fixed by the present because she knows that the universe is governed by exceptionless laws which govern the behaviour of everything in it; thus determinism is often thought of as an empirical hypothesis – one that is tested and confirmed by our experiments on the physical world. This contrasts with fatalism, which is the view held by people who have a (perhaps irrational) belief in the power of fate to bring about an outcome regardless of what we do. For example, in the First World War, there was a belief that if a bullet “had your name on it”, then you were going to get shot, regardless of what precautions you took, or where you went. (For those who have the stomach for it, the same premise is used in the horror film “Final Destination” – teenagers escape an accidental death only for Death to hunt them down, ensuring that they meet a premature end as fate had decreed.)

As ever, Baldrick gets the wrong end of the stick: [http://tinyurl.com/blackadderbullet](http://tinyurl.com/blackadderbullet)
The fatalist says something that the determinist never would: that whatever I do, the end result will be the same. Thus fatalism is a thesis about our powerlessness to influence the future course of events. In contrast, the determinist absolutely believes that our actions really do influence the future course of events: in fact, says the determinist, it is precisely because the universe is governed by fixed, exceptionless laws, that our actions are guaranteed to have an impact – since the laws of nature guarantee that every physical event has consequences, everything we do in the world has an effect, even if we may not be able to predict what that effect is.

That is the reason why determinism and fatalism are incompatible: the fatalist has to imagine some mysterious force in the universe, intervening to block the normal consequences of my actions and make sure that, whatever I do, the end result is the same; how else could a fatalist believe, for example, that the date and time of my death is fixed regardless of whether I live well or badly, and regardless of what precautions I take or risky activities I indulge in?

You should be aware that there is another, rather more respectable, philosophical view which is also sometimes described as “fatalism”; this is logical fatalism, which is a view based on the philosophy of language. The classic argument for logical fatalism is discussed by Aristotle in his De Interpretatione, written in the 4th century BC. It goes like this:

- For every statement \( p \), either it or its negation (“not-\( p \)”)) is true. (This is the “law of excluded middle”)
- So it is either true that there will be a sea battle tomorrow, or it is true that there will not be a sea battle tomorrow.
- So it is already fixed whether there will be a sea battle tomorrow.
- The same argument can be made for any statement about future events.
- So “everything that happens, happens of necessity”.

This argument threatens to show that the future is inevitable – whether or not there will be a sea battle tomorrow is in some sense already decided, and is not something I have the power to bring about or prevent: once it is true that there will be a sea battle tomorrow, that truth-value is fixed and cannot be changed by any action I might choose to perform.

One response to this argument was given by Aristotle himself: that we should reject the first premise, and say instead that it is neither definitely true that there will be a sea battle tomorrow, nor definitely true that there will not be a sea battle tomorrow. This involves endorsing what is sometimes called a “three-valued logic”: instead of two truth-values, “true” and “false”, we should say that statements can have one of three truth-values, “true”, “false”, and “maybe”. Then we can say that the statement “there will be a sea battle tomorrow” is neither true nor false; instead it has a third truth-value which reflects the fact that it is not yet fixed whether the sea battle will occur.

**Determinism, predictability, and subjective probability**

Determinists sometimes say that in principle the future is predictable with absolute certainty: if you were infinitely intelligent, knew all the laws of nature, and knew everything about the state of the universe now, you could predict with absolute certainty what will happen next. If determinism is true, and the combination of the current state of the universe plus the laws of nature leaves us with only one possible future, then it ought to be possible to work out what that future will be given knowledge of the laws and the current state of the universe.
Exercise: how is the description of the workings of the Total Perspective Vortex characteristic of determinism? In what ways does it go beyond what a determinist would accept?

Watch: http://tinyurl.com/TotalPV

The idea that, in a deterministic universe, the future is predictable in principle is associated with Pierre Laplace (1749-1827). He imagined an “intellect” able to grasp and analyse all the data about the current state of the universe and use it to predict the future; this intellect has since come to be known as Laplace’s demon:

“An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.”

Laplace, A Philosophical Essay on Probabilities

Notice that, although the determinist claims that the future is “predictable”, it would be a mistake to argue like this: “I can’t predict the future, therefore determinism is false”. The reason is that the determinist claims only that in principle the future is predictable from the current state of the universe and the laws of nature; but of course we don’t claim to be able to know everything about the current state of the universe, and many scientists believe that we haven’t yet discovered all the laws of nature that govern the behaviour of the universe. Moreover, even if we did know all of this information, working out what was going to happen next would take computing power far in excess of what any modern computer can achieve – you would have to calculate and plot the trajectory of every subatomic particle in the universe!

That’s why it is sometimes pointed out that determinism is compatible with one form of “chance” – namely, subjective probability, or the likelihood we assign to future events based on our own limited knowledge of the world. Of course, you can assign a (subjective) probability to an event even if that event has already been fixed – just as long as you don’t know for certain that the event has been fixed. Here are a couple of examples of how that works:

1. I am waiting for a bus which is ten minutes overdue; as a result of which I decide there is a high (subjective) probability that the bus will arrive in the next five minutes. In fact, it is already fixed or determined that the bus will arrive in the next five minutes (suppose it is just around the corner). Nevertheless – because I don’t know when the bus will arrive – I can still reason about the probability of the bus arriving within a certain time.

2. I toss a coin. Suppose I’m a determinist; in that case I believe that whether it will land heads or tails is already fixed as the coin spins through the air. Nevertheless, I can say that there is a 50% chance of it landing heads, since (from my point of view and limited state of knowledge) either outcome (heads or tails) is equally likely.

The moral of these examples is this: our judgements of probability are based on our own limited knowledge of the world; that’s why we can go on talking about the “probability” of an event happening even if we are determinists and believe that whether the event will happen is already fixed by the current state of the universe and the laws of nature.
Determinism and Predestination

Predestination is the view that God – because he is omniscient (all-knowing) and omnipotent (all-powerful) – already knows, or even has already decided, everything we are going to do throughout our lives. This theological position is associated with John Calvin (1509-1594), and is supported by this passage from the New Testament:

“God is always at work for the good of everyone who loves him. They are the ones God has chosen for his purpose, and he has always known who his chosen ones would be.”

(Romans 8:28)

Theologians have expended a great deal of ink attempting to reconcile the claim that God already knows what we will do in the future, with the claim that our future actions are freely chosen and that we are responsible for them. We’ll pass that by, since for our purposes the most important thing is to understand how determinism – which is the topic of this part of the A-level – is different from the religious notion of predestination. Here are a few important differences:

- Predestination (unlike determinism) does not have to involve a “scientific” conception of the universe as following fixed exceptionless laws of nature. Predestination does not make any claims about how God knows the future – merely that he does; someone might argue that our choices and decisions are not determined by the laws of nature, and yet that God (because he is infinitely wise and knows us infinitely well) still knows what we are going to do in each circumstance we might encounter.

- If we interpret the supporter of predestination as saying that God has already decided which of us will go to heaven and which go to hell, as Calvin believed, then we might imagine God intervening in our lives to ensure that certain events happen to us, or even to ensure that we make certain decisions. This would involve God intervening to block the normal consequences of our actions – in effect, putting the normal laws of nature on hold so that he could bring about a “miracle” to change the course of our lives. So predestination allows the possibility of God intervening to block the normal course of events, while determinism implies that actions always have the result predicted by the laws of nature which govern the universe.

- While predestination can be true only if God exists, the truth of determinism does not entail any conclusion about the existence or non-existence of God: you might believe in a God who brought the universe into existence along with all the deterministic laws, then simply left it running like a complex clockwork machine, having already worked out what was going to happen as a consequence of how the universe was set up at the start. (This view is known as deism to contrast it with theism, which is the belief in a God who actively intervenes in the world.)
3 – Arguments for determinism

Aims: to understand different reasons for believing in determinism; to understand some important criticisms of determinism.

First, a clarification: we’ve defined determinism as a claim about the total state of the universe: that the current state of the universe, plus the laws of nature, fixes everything that is going to happen next. However, you should be aware that people often argue for determinism by trying to demonstrate the lesser claim that human actions are determined or “fixed in advance” by circumstances. The assumption here is that it is uncontroversial that the behaviour of ordinary physical objects is fixed in advance by the laws of nature – but we need some extra reason for thinking that the behaviour of people is also subject to laws which fix what we are going to do next based on how things are now. By showing that the behaviour of people is subject to deterministic laws, we go a long way to establishing the determinist’s claim that everything that happens in the universe is determined in advance. The behaviour of people is the most obvious counter-example to such a universal claim, so by showing that this behaviour is also governed by deterministic laws, we defend universal determinism by rejecting a counter-example to it.

Determinism and human nature

One way of arguing that human actions are determined or “fixed in advance” is to point out that we are already committed to the view that human behaviour is predictable; and this predictability seems to presuppose that people’s behaviour is fixed in advance by their character and the situations they find themselves in. Hume offers that kind of argument, suggesting that everyone (if they think about it enough) should agree that human nature is determined in advance:

“It is universally acknowledged that there is a great uniformity among the actions of men, in all nations and ages, and that human nature remains still the same, in its principles and operations. The same motives always produce the same actions. The same events follow from the same causes. Ambition, avarice, self-love, vanity, friendship, generosity, public spirit: these passions, mixed in various degrees, and distributed through society, have been, from the beginning of the world, and still are, the source of all the actions and enterprises, which have ever been observed among mankind. Would you know the sentiments, inclinations, and course of life of the Greeks and Romans? Study well the temper and actions of the French and English: You cannot be much mistaken in transferring to the former most of the observations which you have made with regard to the latter. Mankind are so much the same, in all times and places, that history informs us of nothing new or strange in this particular. Its chief use is only to discover the constant and universal principles of human nature, by showing men in all varieties of circumstances and situations, and furnishing us with materials from which we may form our observations and become acquainted with the regular springs of human action and behaviour.

…”

Hence likewise the benefit of that experience, acquired by long life and a variety of business and company, in order to instruct us in the principles of human nature, and regulate our future conduct, as well as speculation. By means of this guide, we mount up to the knowledge of men’s inclinations and motives, from their actions, expressions, and even gestures; and again descend to the interpretation of their actions from our knowledge of their motives and inclinations. The general observations treasured up by a course of experience, give us the clue of human nature, and teach us to unravel all its intricacies. Pretexts and appearances no longer deceive us. Public declarations pass for the specious
colouring of a cause. And though virtue and honour be allowed their proper weight and authority, that perfect disinterestedness, so often pretended to, is never expected in multitudes and parties; seldom in their leaders; and scarcely even in individuals of any rank or station. But were there no uniformity in human actions, and were every experiment which we could form of this kind irregular and anomalous, it were impossible to collect any general observations concerning mankind; and no experience, however accurately digested by reflection, would ever serve to any purpose. Why is the aged husbandman more skilful in his calling than the young beginner but because there is a certain uniformity in the operation of the sun, rain, and earth towards the production of vegetables; and experience teaches the old practitioner the rules by which this operation is governed and directed.

66. We must not, however, expect that this uniformity of human actions should be carried to such a length as that all men, in the same circumstances, will always act precisely in the same manner, without making any allowance for the diversity of characters, prejudices, and opinions. Such a uniformity in every particular, is found in no part of nature. On the contrary, from observing the variety of conduct in different men, we are enabled to form a greater variety of maxims, which still suppose a degree of uniformity and regularity.

... 67. I grant it possible to find some actions, which seem to have no regular connexion with any known motives, and are exceptions to all the measures of conduct which have ever been established for the government of men. But if we would willingly know what judgement should be formed of such irregular and extraordinary actions, we may consider the sentiments commonly entertained with regard to those irregular events which appear in the course of nature, and the operations of external objects. All causes are not conjoined to their usual effects with like uniformity.

... Thus, for instance, in the human body, when the usual symptoms of health or sickness disappoint our expectation; when medicines operate not with their wonted powers; when irregular events follow from any particular cause; the philosopher and physician are not surprised at the matter, nor are ever tempted to deny, in general, the necessity and uniformity of those principles by which the animal economy is conducted. They know that a human body is a mighty complicated machine: That many secret powers lurk in it, which are altogether beyond our comprehension: That to us it must often appear very uncertain in its operations: And that therefore the irregular events, which outwardly discover themselves, can be no proof that the laws of nature are not observed with the greatest regularity in its internal operations and government.

68. The philosopher, if he be consistent, must apply the same reasoning to the actions and volitions of intelligent agents. The most irregular and unexpected resolutions of men may frequently be accounted for by those who know every particular circumstance of their character and situation. A person of an obliging disposition gives a peevish answer: But he has the toothache, or has not dined. A stupid fellow discovers an uncommon alacrity in his carriage: But he has met with a sudden piece of good fortune. Or even when an action, as sometimes happens, cannot be particularly accounted for, either by the person himself or by others; we know, in general, that the characters of men are, to a certain degree, inconstant and irregular. This is, in a manner, the constant character of human nature; though it be
applicable, in a more particular manner, to some persons who have no fixed rule for their conduct, but proceed in a continued course of caprice and inconstancy. The internal principles and motives may operate in a uniform manner, notwithstanding these seeming irregularities; in the same manner as the winds, rain, clouds, and other variations of the weather are supposed to be governed by steady principles; though not easily discoverable by human sagacity and enquiry.

69. Thus it appears, not only that the conjunction between motives and voluntary actions is as regular and uniform as that between the cause and effect in any part of nature; but also that this regular conjunction has been universally acknowledged among mankind, and has never been the subject of dispute, either in philosophy or common life.”

Hume, *An Enquiry concerning Human Understanding*, VIII.65-9

*Question:* In your own words, explain what Hume’s response is to someone who argues that human behaviour is *not* determined, because it is not the case that “all men, in the same circumstances, will always act in precisely the same manner”.

Notice that Hume does not concern himself with the question of whether the universe as a whole is governed by deterministic laws; his aim is simply to show that human actions should be viewed as determined by “constant and universal principles” which fix what each person will do next, given what has happened so far.

A modern counterpart of Hume’s view can be found in the “nature v. nurture” debate. Whenever someone commits a terrible crime, the media reopen a debate about whether the criminal’s behaviour is the result of hereditary and genetic factors (“nature”), or of the environment they were brought up in (“nurture”). The presupposition of such debates is that all human action must be the inevitable result of either environmental or hereditary factors, and all we need to do is to establish which of these is more important in this specific case. Similarly, scientists know that one way of ensuring that their research will make headlines is to claim to have found “the gene for X”, where “X” could be obesity, alcoholism, or even entrepreneurship:
**Genes key to entrepreneurs' drive** (news.bbc.co.uk, 5/6/2006)

*Entrepreneurs are largely born rather than made, research suggests.*

A UK-US study has found our genes are crucial in determining whether we are entrepreneurial and likely to become self-employed. It found nearly half of an individual's propensity to become self-employed is due to genetic factors. And, contrary to previous beliefs, family environment and upbringing have little influence on whether a person becomes self-employed or not.

The research is important for business schools and employers who in the future could identify ways of selecting those who were most likely to succeed. The other factors which did play a significant role were random life events, such as being made redundant, winning a large sum of money, or a chance meeting.

The study was carried out by the Twin Research Unit at St Thomas' Hospital, London, the Tanaka School of Business at Imperial College, London and the US Case Western Reserve University. The researchers examined self-employment in 609 pairs of identical twins and 657 pairs of same-sex non-identical twins in the UK. Identical twins share all their genes while non-identical twins share, on average, about half. The rate of entrepreneurship among twins was the same as across the general population.

But researchers looked at whether one twin being an entrepreneur increased the chance of their co-twin becoming an entrepreneur. By comparing the difference in similarity rates between identical and non-identical twins they are able to establish the importance of genetic and environmental factors. The similarity rate within the identical twins group was greater than for the non-identical twin group which suggests that genes are important.

Professor Tim Spector, director of the Twin Research Unit, said:

"This relatively high heritability suggests the importance of considering genetic factors to explain why some people are entrepreneurial, while others are not. The research is important for business schools and employers who in the future could identify ways of selecting those who were most likely to succeed."

Professor Spector said there was evidence to show that genetic factors influence a variety of business-related areas, from job satisfaction to vocational interests.

For discussion: do you agree that everything that a person does in her life is the result either of environmental or of hereditary factors? Is there any good reason to believe that this must be true?
Social, Historical, and Economic Determinism

Those who believe that our behaviour is fixed by our genetic inheritance are often called “genetic determinists”; similarly, we could use “social determinism” as a term for the view that our behaviour is fixed by our upbringing and social conditioning. Another important theory that you should be aware of is the economic determinism associated with Karl Marx (1818-83) and his theory of historical materialism. Marx argued that all human action is the result of the need to engage in certain kinds of economic activity to satisfy our needs – and that the sort of person we are is the result of the kind of activity we engage in. In Marx’s terminology, we take on a “mode of life” associated with the kind of job we have, and this mode of life determines how we behave and interact with one another. For example, someone who works as a school teacher to pay for food and housing will develop a character which is the direct consequence of, and determined by, the specific economic situation he finds himself in.

“In the social production of their existence, men inevitably enter into definite relations, which are independent of their will, namely relations of production appropriate to a given stage in the development of their material forces of production.

The totality of these relations of production constitutes the economic structure of society ... The mode of production of material life conditions the general process of social, political, and intellectual life.

It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness.”

Moreover, Marx seems to suggest that it is not just the characteristics of individuals that are determined or “conditioned” by their economic circumstances; he also suggests that the religious, artistic, moral and philosophical beliefs within a society (which he called the society’s “ideology”), and even the historical and political events which take place within a society, are all determined by the economic circumstances which take obtain within the society. Thus Marx is sometimes interpreted as claiming that the path of history is not only predictable but even inevitable: all human societies are determined to move from feudalism to capitalism to communism as a result of changing economic structures.

Caveat: the interpretation of Marx’s own views is contentious; in particular it is not clear that Marx himself would have agreed with the view expressed by self-identified “Marxists” such as Lenin, that every society must eventually become communist.
Human action and science

We have already met one important reason for believing that everything we do is determined; this is that human action seems as though it must be subject to the laws of nature just as much as any other events. The reason is simple: every human action involves some kind of change in the molecules that make up our body – for example neurons firing to control the movement of our muscles – and there is no reason for thinking that the molecules that make up our body are somehow exempt from the laws of nature that fix the behaviour of other molecules in the universe. Surely electrical activity in the brain is governed by the same laws that govern electrical activity everywhere in the world; and in that case surely whatever happens in my brain is fixed in advance by what has happened already plus the laws of nature.

Determinism and Causation

Could determinism be known to be true a priori? You will remember that Kant believed we know a priori that every event has a cause “from which it follows according to a rule”; this is the thesis of universal causation. Causes are generally supposed to necessitate their effects, so to say that every event has a cause is to say that every event is necessitated or determined by some previous event or circumstance – given the cause happened, the effect had to happen. For example, the event of my striking the match caused it to ignite – which is simply to say that, given that I stuck the match in exactly this way, the laws of nature ensured that the match would catch fire; there just wasn’t anything else that could have happened as a consequence of this particular event in these particular circumstances. So if we believe that every event has a cause which necessitates it, this amounts to a belief in a deterministic universe – one where everything that happens is necessitated by something that has already happened.

Of course, you might reject the idea of universal causation: if you believe that the universe had a first event (e.g. the “Big Bang”) then there must be at least one event which was not caused by a preceding event. Alternatively, you might say that we do not know a priori that every event has a cause; instead universal causation – and determinism itself – is an empirical hypothesis: i.e. a hypothesis reached by inductive generalization from the fact that every event we have encounter so far seems to have a cause, and the fact that the behaviour of everything scientists have so far observed in the world seems to be governed by deterministic laws of nature. If that is so, then determinism can never be proved with absolute certainty to be true – for there is no way of observing every single particle in the universe to check that it is obeying the same laws as the rest. There is no way to rule out the hypothesis that the universe contains some events which “just happen” without a preceding cause, or to prove that there are never exceptions to the laws of nature which the scientists try to discover.

Criticism of determinism

One problem for the determinist is that her claim “everything is determined” outstrips the available evidence; we are really only entitled to say that the events we have observed so far appear to be governed by deterministic laws, but not to argue that everything that happens must be fixed in advance. On the other hand, if we reject the suggestion that determinism is established on the basis of empirical observation, and instead try to claim that we know a priori that determinism is true, we need to explain why we are so sure that our belief that everything is determined is a genuine insight into the way the world is, rather than simply an expression of our own closed-mindedness and prejudice.

Similar criticisms had already been offered by Hume: he pointed out that we can never observe one event necessitating another; instead we simply see one thing happening, followed by another. So our belief that events in the universe are necessitated by other events is simply an expression of our own expectation that
similar causes will have similar effects; this expectation cannot be **rationally justified**, but is just a result of our natural tendency to assume a “necessary connexion” where really all we can observe is a “constant conjunction”.

Scientists might argue that the success of science in making predictions about what is going to happen next can justify our faith that *everything* in the universe is governed by deterministic laws, but Hume had an answer to this as well: the fact that events we have observed in the past were predictable according to deterministic laws does not give us the right to be certain that events in the future will also be predictable according to such laws – just as the fact that the sun has risen when we expected it to on every day so far does not rule out the possibility that the sun will not rise tomorrow.

Another problem for determinism is that many scientists now believe that it is not true that the tiniest physical particles obey deterministic laws; instead, they claim, quantum mechanics shows us that events at the level of subatomic particles are not completely determined by preceding events plus the laws of nature. The idea is that, at the subatomic level, there are genuinely random or “chancy” events. Although what has happened already might influence what happens next by raising or lowering the probability of such an event occurring, it is not the case that the event is determined or necessitated by what has happened already, since there is always a chance that the event will not occur as well as a chance that it will occur. Supporters of quantum mechanics thus reject the claim that laws of nature should enable us in principle to predict the future with absolute certainty; all our knowledge of laws of nature can do is enable us to calculate the probability of an event happening given what has happened already.

Another **caveat** is in order here: neither scientists nor philosophers have been able to agree about the correct interpretation of quantum mechanics. On some interpretations, the “randomness” of events at the subatomic level shows us that the universe itself is objectively random and not governed by deterministic laws; however, on other interpretations the apparent “randomness” is just a reflection of the limitations of our own knowledge — quantum mechanics as it stands is “incomplete”, and reality contains some other foundational property or “hidden variable” which determines what happens in any given circumstance. On these interpretations, quantum mechanics is an incomplete description of a deterministic universe governed by laws we have not yet fully understood. The most famous supporter of such deterministic interpretations of quantum mechanics was Albert Einstein himself, who remained convinced throughout his life that there could not be such a thing as objectively random events — or as he put it, “God does not play dice with the universe”.
4 – Incompatibilism I: Determinism as undermining free will

**Aims:** to understand what is meant by “incompatibilism”; to understand what “hard determinism” is and why it seems plausible. To understand how the experience of Free Will might be an illusion; to understand the idea that rational explanation and causal explanation are incompatible.

**“Incompatibilism” and “Hard Determinism”**

Perhaps the most basic question within this topic is, is Free Will compatible with determinism? This is the question whether the two claims (i) that humans have Free Will, and (ii) that determinism is true, can both be true at the same time. In other words, is Free Will possible in a deterministic universe where everything that happens is fixed in advance by what has happened already plus the laws of nature? Perhaps unsurprisingly, philosophers who claim that Free Will and determinism are compatible are called “compatibilists”, while those who deny that they are compatible are called “incompatibilists”.

You should notice that there are two ways of being an incompatibilist. First, you might reason like this: Free Will is incompatible with determinism, and we have good reason to believe that determinism is true; therefore we do not have Free Will. This position is known as hard determinism; it is committed both to the truth of determinism, and to the claim that we do not have Free Will.

Alternatively, you might reason like this: Free Will is incompatible with determinism, and we have good reason to believe that we have Free Will; therefore determinism is false. This position is known as libertarianism; we’ll look at it in more detail next week. For now the important thing is to understand the arguments which persuade both hard determinists and libertarians that Free Will is incompatible with determinism.

**The Argument from Alternative Possibilities**

Determinists believe that the current state of the universe, plus the laws of nature, fixes what is going to happen next: given the way things are at a certain time, there is only one possible way things could turn out next, namely the way things actually turn out. So although it seems to you like there are many possible different ways you could have acted in a given situation, there was really no other way you could have acted; the deterministic laws of nature fixed that there was only one possible outcome of the situation you were in.

How does this interfere with Free Will? To say that we have Free Will is to say that we can make free choices, and to make genuinely free choices, it seems we must be able to choose between alternative courses of action that are open to us. But if determinism is true, then there are no possible alternative courses of action open to us – we can only “choose” the option that is already fixed for us by the current state of the universe plus the laws of nature, since it is literally impossible for us to choose anything else. And of course, a “choice” where only one option is possible is no choice at all.

This argument is sometimes summed up by saying that Free Will requires that we “could have done otherwise” in the circumstances – and of course, since determinism says that, in the circumstances, there was only one way things could have turned out, it seems that the determinist must deny that we “could have done otherwise” and hence deny that we have Free Will.

nb: Compatibilists (who reject this argument, and claim that Free Will and determinism are compatible) usually respond by saying that the argument makes a mistake about what is required for an action or decision to be truly “free”; they then have to come up with an account of what is meant by “free choice”
which does not require that a “free” choice involves the possibility that the agent could have chosen differently in the circumstances.

The Consequence Argument

An alternative argument for the incompatibility of Free Will and determinism was already known to ancient philosophers, and revived in the 1960s by Carl Ginet. Here’s a version of the argument:

- We have Free Will only if our actions are under our own control.
- Our actions are consequences of (determined by) the past and the laws of nature.
- We have no control over the past or over the laws of nature.
- Therefore, we have no control over our actions.
- Therefore, we do not have Free Will.

The central idea is a simple one: if our actions are consequences of things we have no control over (the past and the laws of nature) then we cannot claim to have control over our actions: our actions are not “up to us”, because they were already fixed by something that was not up to us, namely the state of the universe at some time in the distant past plus the laws of nature. (Remember that the determinist claims that the total state of the universe at any time, plus the laws of nature, determines what the state of the universe will be at any subsequent time – so our actions were already fixed by the state of the universe at the time of the dinosaurs, say.)

One way a compatibilist might challenge this argument would be to suggest that we can still “control” our actions even if those actions were already fixed by the state of the universe in the distant past: I have control over my actions just so long as those actions were caused by my beliefs and desires, and it doesn’t matter that those beliefs and desires are part of a causal chain stretching far into the past, over which I have no control. We would then “control” our actions in the same sense that a thermostat “controls” the temperature in a house, even though what the thermostat does is controlled in turn by the temperature readings its sensor picks up, and the temperature it is set to by the owner of the house:

“A thermostat can control the future, in respect of temperature. That is what thermostats do. A thermostat controls the future by being part of the way in which the past controls the present and future. And according to compatibilism, that is the way we control things. We are involved in the causal order. We are part of the way in which the past controls the present and future.”

Blackburn, Think, p.44

The compatibilist will say that it doesn’t matter what caused us to have the beliefs and desires we have; what matters for us to have control over our actions is only that our actions are caused by our beliefs and desires rather than something else.

The Experience of Free Will as an Illusion

If you are persuaded by these arguments for incompatibilism, and you are convinced that the universe is deterministic, then the rational thing to do is to become a hard determinist, and deny that human beings have Free Will. Then, of course, you need to explain how it is that it seems to us that we are making free choices and performing free actions, given that you are convinced that these actions are not free. The experience of Free Will is most strongly apparent when we are deliberating about possible courses of action, or trying to “make up our minds” whether to do something. Here it seems that there is more than one possible course of action open to us, and we are freely deciding which of these possibilities to act on.

As Peter Van Inwagen notes, this seems to presuppose the existence of genuine alternative possibilities:
“I would ask you to try a simple experiment. Consider some important choice that confronts you. You must, perhaps, decide whether to marry a certain person, or whether to undergo a dangerous but promising course of medical treatment, or whether to report to a superior a colleague you suspect of embezzling money. (Tailor the example to your own life.) Consider the two courses of action that confront you; since I don’t know what you have chosen, I’ll call them simply A and B. Do you really not believe that you are able to do A and able to do B? If you do not, then how can it be that you are trying to decide which of them to do? It seems clear to me that when I am trying to decide which of two things to do, I commit myself, by the very act of attempting to decide between the two, to the thesis that I am able to do each of them. If I am trying to decide whether to report my colleague, then, by the very act of trying to reach a decision about this matter, I commit myself both to the thesis that I am able to report him and to the thesis that I am able to refrain from reporting him: although I obviously cannot do both these things, I can (I believe) do either. In sum: whether we are free or not, we believe that we are.”

Van Inwagen, The Mystery of Metaphysical Freedom

One possible response the hard determinist might make is this: science tells us that the actual process that goes on when we make a decision is much more complicated than simply weighing option A against option B; instead, it involves the firing of billions of neurons in our brain in very complicated patterns. We don’t know which decision we’re going to make in advance because we don’t have sufficient knowledge of the inner workings of our brains to work out what those neurons are going to do. But that doesn’t show that the eventual decision isn’t determined in advance—our failure to predict our own decision in advance simply reflects our lack of knowledge of our own brains, rather than showing that our decisions are undetermined. This response is simply a specific instance of something we have met already: the fact that we cannot use our inability to predict an outcome to show that the outcome is not determined—because our inability to make the prediction might reflect our lack of knowledge of the factors determining the outcome rather than revealing that there is a genuine, objective chance that things might turn out differently.

However, this response can be criticized for its excessively reductionist approach to human minds and rationality—it seems to suggest that “making a decision” can be “reduced to” nothing more than the firing of certain neurons in the brain, and this kind of reductive claim has been attacked by philosophers because it embodies an overly-simplistic view of the relationship between mind and brain. In particular, this kind of account of decision-making seems to undermine human rationality, since it tries to replace concepts such as reasons, desires, and decisions with the scientific vocabulary of neurons firing. A scientific description of the brain has no room for concepts like a “reason”—so if we commit ourselves to the view that there is nothing more to deciding to act than the firing of neurons in the brain, we risk losing sight of what it is that makes us rational agents rather than just very complex machines. This point is sometimes put as the claim that rational explanation and causal explanation are incompatible: we can explain an action in terms of the physical causes that made it happen, or in terms of the reasons which led the agent to choose to do it; but we can’t do both at the same time—or at least, rational and causal explanations belong to different “stances” we might take to the same creature, and we can only take one “stance” at a time.

Can the determinist find a way to explain away the appearance of free choice without implying that human minds are nothing more than brains? One way might be to think more carefully about what really happens when we make a decision. We usually think of ourselves as having desires of different strengths, and weighing up courses of action on the basis of how likely they are to result in the fulfilment of our desires. Sometimes we might not know which course of action we value the most until we choose it. So “making a
decision” is often a way of finding out what it is that we want most, or which reason is more important to us. And if we endorse that way of thinking about reasoning, then it is a short step to the view that what seems like a free, undetermined choice is in fact completely determined in advance by the relative strengths of the reasons and desires on either side of the argument – even though we only find out what we want the most when we make the decision. So again, our inability to predict a decision before we make it does not show that the decision is not fixed in advance – it merely reflects that we cannot always tell which reasons and desires are more important to us before we compare them in deliberation.

The 19th-century German philosopher Schopenhauer vividly illustrates how it might be possible for us to believe that we are making free choices when in fact we are not: all we have to do to feel like we are free is to imagine ourselves doing things other than we actually do, even if those are things that we (really) never could do.

“Let us imagine a man who, while standing on the street, would say to himself: ‘It is six o’clock in the evening, the working day is over. Now I can go for a walk, or I can go to the club; I can also climb up the tower to see the sun set; I can go to the theatre; I can visit this friend or that one; indeed, I also can run out of the gate, into the wide world and never return. All this is strictly up to me; in this I have complete freedom. But still, I shall do none of these things now, but with just as free a will I shall go home to my wife.’ This is exactly as if water spoke to itself: ‘I can make high waves (yes! in the sea during a storm), I can rush down hill (yes! in the river bed), I can plunge down foaming and gushing (yes! in the fountain) I can, finally, boil away and disappear (yes! at certain temperature); but I am doing none of these things now, and am voluntarily remaining quiet and clear in the reflecting pond.’” Schopenhauer, On the Freedom of The Will

Exercise: make two diagrams – (i) showing the relationship between compatibilism, incompatibilism, determinism, hard determinism, and libertarianism; (ii) illustrating the two arguments for incompatibilism covered in this section.
5 – Incompatibilism II: Defending Free Will

Libertarianism

Libertarians – like hard determinists – claim that Free Will and determinism are incompatible; however, they approach the debate from the other direction, arguing that our immediate experience of free choice is more certain than our belief in determinism: because we have good reason to believe that we have Free Will, we therefore have good reason to believe that determinism is false, and to believe that some events in the world (namely, our own free choices) are not fixed in advance by what has happened already plus the laws of nature. Here we’ll consider two kinds of libertarian theory: one represents the human mind as something which stands outside the natural causal order; while the other identifies our freedom with random or uncaused quantum events in the brain.

The mind as allowing human decision-making to occupy a special place outside of the natural order.

Although the view is not as popular as it used to be, many philosophers have suggested that the mind is a fundamentally different kind of thing from the brain; although the mind uses the brain to control the body, the mind itself is not a physical object, and so is not subject to the deterministic laws of nature which govern the behaviour of physical objects. And if minds are not subject to laws of nature, then the behaviour of minds is not determined in advance in the same way that the behaviour of physical objects is determined in advance. So while the behaviour of physical objects is determined in advance, our decisions are not, since we are not physical objects.

This view – Cartesian dualism, that the mind is a different kind of thing from the brain – was elaborated and defended by Descartes. His main reason for believing it was that he could conceive of (imagine) himself without a body, but could not imagine himself without a mind; since mind and body are separable in this way in thought, he concluded that they must also be “separate” or “distinct” things in reality. This gives us a way of seeing how Free Will might be possible: all of the actions associated with making decisions or “willing” things to happen (e.g. understanding, deliberating, deciding, planning) are mental activities, so showing how mental activity is not subject to determinism should allow us to understand how our decisions can count as free.

However, this kind of libertarian response to the Free Will problem faces some significant objections. First, there is the interaction problem: how is it possible for our decisions or “volitions” (acts of will) to cause physical effects; how can something non-physical (the mind) control on or interact with something in the physical world, namely our brain? The problem seems especially pressing when we remember that, according to the Cartesian dualist, minds are immaterial things and so do not even have a physical location. How can something that lacks a physical location interact with something located in the world? The idea of such interaction seems impossible to make sense of, or “incoherent”.

Even if we find a way to answer the interaction problem, another issue threatens: we originally endorsed the dualist theory to show how minds could make free choices even though physical objects seem to be subject to deterministic laws which fix their behaviour in advance. But what about the physical object that is my body? If physical bodies behave deterministically then their movements are fixed in advance by the previous state of the universe plus the physical laws of nature, regardless of what our minds decide to do, and it looks as if our choices and “volitions” are impotent (powerless) to affect what our bodies actually do. (nb: some philosophers have defended this kind of view, which is known as epiphenomenalism: it is the view that our conscious minds have no influence on what our bodies actually do; we observe what is going...
(on without influencing it, like a toddler with a toy steering wheel who genuinely believes he is driving the
car.) On the other hand, if we decide that our minds are capable of causally influencing the physical world
(e.g. by causing particular neurons to fire in our brains), then these events in our brain would not have a
prior physical cause, and could not be explained by reference to the normal laws of nature governing the
behaviour of the physical world. So it seems that this kind of dualist libertarianism requires us to accept
that there are physical events which do not fall under physical laws of nature; we cannot defend Free Will
by arguing that bodies are, and minds are not, subject to deterministic laws, since this version of
libertarianism requires that there are also some physical events that are not determined in advance.

Finally, we should notice that this view will struggle to do justice to the way in which the mind depends
on different parts of the brain for its different functions. We know that damage to specific parts of the brain
can deprive us of specific mental functions – for example, injury to part of the left frontal lobe of the brain
can result in aphasia, the inability to produce language. Rather than saying that the mind controls the brain,
it seems more correct to say that our mental functions are carried out by different parts of the brain; and it
is the brain that is ultimately responsible for mental acts such as willing and deciding, these actions will be
ultimately subject to deterministic physical laws, and so are not free.

**Exercise:** using the internet, research mind-brain dependence. How many different mental functions can be
identified as carried out by specific regions of the brain?

**Libertarianism as requiring a gap in the causal order of the physical world**

The previous kind of libertarianism attempted to respect the view that physical events are governed by
strict, deterministic laws while saying that our choices and decisions are not determined, because they are
the actions of a non-physical and separate mind. An alternative way of arguing for libertarianism is to adopt
the suggestion made by some interpretations of quantum mechanics, that the world contains genuinely
“random” events at the microphysical level of atomic and sub-atomic particles; events which are not
determined in advance by laws of nature. Because these random quantum events must also be occurring in
our brains, we might be able to use them to argue that it is not the case that everything we do is fixed in
advance: if our decision about what to have for dinner depends on a microphysical quantum event in the
brain that is not determined in advance by the laws of physics, then there will be a genuine chance that we
could decide differently right up until the point that we make our decision. This view has been defended
recently by Robert Kane:

> “Imagine that ... efforts of will ... are complex chaotic processes in the brain, involving neural
networks that are globally sensitive to quantum indeterminacies at the neuronal level. Persons
experience these complex processes phenomenologically as “efforts of will” they are making to
resist temptation in moral and prudential situations. These efforts are provoked by the competing
motives and conflicts within the wills of the persons.” Kane, *The Significance of Free Will* (1998),
p.130

Kane’s point is that the brain somehow “magnifies” tiny quantum indeterminacies – randomness at the
sub-atomic level – to the point where there are genuine alternative possibilities about how we will act; the
existence of indeterminism at the quantum level is reflected by indeterminism at the level of our decisions
and actions. So our actions are not determined by the previous state of the universe and the laws of nature,
and so there is no reason to believe that our choices are not free. Interestingly, the basis of this idea is as
old as the ancient Greeks: the ancient philosopher Epicurus was the first to suggest that our free choices
are the result of “swerves” in the motion of atoms. These “swerves” are supposed to happen at random,
and enable us to say that there was always a chance that we could not have done what we did – as there was always a chance that a random Epicurean “swerve” in the brain would make us do something different.

This kind of appeal to indeterminism faces a serious objection, however: suppose that it is true that my decisions are not fixed in advance by deterministic laws of nature, but instead depend on random quantum events at the subatomic level in my brain. So my decisions ultimately depend on the random behaviour of tiny particles in my brain. How then can I claim that my decisions are up to me? I don’t have any control over the quantum events in my brain; so – if those quantum events control my decisions – it seems that I don’t have any control over my decisions. This objection is raised by Van Inwagen:

“Does postulating or asserting that the laws of nature are indeterministic provide any comfort to those who would like to believe in metaphysical freedom? If the laws are indeterministic, then more than one future is indeed consistent with those laws and the actual past and present—but how can anyone have any choice about which of these futures becomes actual? Isn’t it just a matter of chance which becomes actual? If God were to “return” an indeterministic world to precisely its state at some time in the past, and then let the world go forward again, things might indeed happen differently the “second” time. But then, if the world is indeterministic, isn’t it just a matter of chance how things did happen in the one, actual course of events? And if what we do is just a matter of chance—well, who would want to call that freedom?

It seems, therefore, that, in addition to our argument for the incompatibility of metaphysical freedom and determinism, we have an argument for the incompatibility of metaphysical freedom and indeterminism.”

Van Inwagen, *The Mystery of Metaphysical Freedom*.

Moreover, as we’ve already noted, not all scientists agree that the apparent “randomness” of quantum events reflects genuine randomness in the way the world is – this “randomness” might be a defect of our theory rather than a feature of the world itself.

**Exercise:** listen to the BBC “In Our Time” programme on Free Will at [http://tinyurl.com/IOTfreewill](http://tinyurl.com/IOTfreewill), and answer the following questions:

1. What, according to Simon Blackburn, is meant by “free will”?
2. What did Hume think was “the only freedom he needed”?
3. What two reasons does Simon Blackburn give for philosophers’ interest in the topic of Free Will?
4. What everyday example does Helen Beebee give to illustrate determinism?
5. How does she illustrate the difference between determinism and fatalism, or belief in “fate”?
6. What aspect of religion is said to be hard to combine with determinism?
7. How is the “most commonly accepted version of Free Will” characterized?
8. What, according to Simon Blackburn, is the correct account of moral responsibility?
9. In one sentence, what is the point of the example about being offered £100 to torture someone?
10. Why, according to Galen Strawson, isn’t it going to help if determinism is false?
11. What is Galen Strawson’s four-stage argument for pessimism?
12. What is the point of the example of negligence when looking after a child?
13. Why does Helen Beebee reject libertarianism?
14. What was David Hume’s definition of “liberty”?
15. What did Peter Strawson argue to try to subvert the debate about Free Will?
6 - Compatibilism/soft determinism: free will as compatible with determinism.

Aims: to understand how “free” or “voluntary” action can be causally determined and yet distinguishable from psychologically or physically constrained action; to understand how voluntary action can be defined in terms of the type of cause from which it issues. To understand how “guidance control” or “second-order desires” can feature in compatibilist definitions of freedom.

So far we have covered two approaches to the debate: (i) rejecting Free Will (and keeping determinism); (ii) rejecting determinism (and keeping Free Will). There is, of course, a third way of responding: to find a way of somehow arguing that Free Will and determinism are compatible – i.e. that it is possible for determinism and Free Will to coexist; that it can be true both that our actions are determined by the laws of nature and what has already happened and that our actions are free. Philosophers who take this position are called (unsurprisingly) compatibilists. They are also sometimes known as soft determinists – “determinists” because they accept that determinism is (or is likely to be) true; “soft” because they do not think the truth of determinism prevents us from being free.

Any compatibilist approach to the Free Will problem is going to have to find a way to respond to the arguments for incompatibilism covered in week 4; the usual way compatibilists try to do this is to show that incompatibilist arguments rest on a mistake about what freedom really is. This might sound like the debate becomes merely about the definition of the word “freedom” – in which case the argument might seem trivial; however, you could also think of this as an exercise in conceptual analysis – investigating and clarifying a concept of freedom which everyone shares but which most people do not fully understand. (nb: in the 20th century many philosophers believed that all philosophy consisted of “conceptual analysis”. Fortunately this intellectual trend is less fashionable than it used to be.)

Case study: David Hume

David Hume was one of the first philosophers to suggest that determinism and Free Will are compatible:

“For what is meant by liberty, when applied to voluntary actions? We cannot surely mean that actions have so little connexion with motives, inclinations, and circumstances, that one does not follow with a certain degree of uniformity from the other, and that one affords no inference by which we can conclude the existence of the other. For these are plain and acknowledged matters of fact. By liberty, then, we can only mean a power of acting or not acting, according to the determinations of the will; this is, if we choose to remain at rest, we may; if we choose to move, we also may. Now this hypothetical liberty is universally allowed to belong to every one who is not a prisoner and in chains. Here, then, is no subject of dispute.”

Hume, Enquiry VIII.i

Hume, then, disagrees with those who say that a free action is one that “could have been otherwise” in the given situation: for an action to be my action, it must “follow with a certain degree of uniformity” from my “motives, inclinations, and circumstances”. Elsewhere (in the Treatise on Human Nature) Hume puts this point by saying that, for an action to be my action it must be caused by my desires and circumstances – which is as good as saying that any action of mine must be necessitated by my desires and circumstances.

Hume’s proposed alternative definition of “liberty” or “Free Will” needs some explanation. He describes freedom as “a power of acting or not acting, according to the determinations of the will”. By this he emphatically does not mean that in every situation there must be a possibility either that we will do something or that we will not do it; he means only that we are free so long as we have the power to do
what we want to do. So we are free to stand up, because this is something it is in our power to do if we want to. Of course, if determinism is true, then what we want to do (and what we actually do) is fixed in advance. But that doesn’t stop us from being free in Hume’s sense: according to him it doesn’t matter that our desires are fixed in advance by circumstances outside our control; all that matters is that there is nothing preventing us from acting according to our desires.

For Hume, the difference between a “free” or “voluntary” action and an involuntary one is the **type of cause** from which it issues: voluntary actions are those that are caused by my desires and beliefs, while involuntary actions are those which are caused by something else. If my action of kicking the doctor in the face was caused by a reflex action as he struck my patellar tendon with a hammer, then my action was involuntary; however, if my action was caused by my deep personal dislike of the doctor and my desire to hurt him, then my action was voluntary. The crucial point here is that both involuntary and voluntary actions are caused, and hence necessitated; the difference between them is merely what they are caused by. So Hume thinks of Free Will as *causally determined and yet distinguishable from psychologically or physically constrained action*: an action which is causally determined by my desires is voluntary and “free”, while an action which is “constrained” or restricted by physical circumstances or some psychological problem (such as the inability to act on my desires through depression or extreme terror) is involuntary and hence not free in Hume’s sense.

**Criticisms of Hume**

The most famous critic of Hume’s account was Immanuel Kant, who described this approach to the Free Will problem as a “wretched subterfuge”, claiming that the kind of freedom that Hume attributes to people could also be attributed to a clock or a projectile flying through the air, since in each case the motion is not down to “anything external”:

“Some try to evade this [denial of Free Will] by saying that the causes that determine his causality are of such a kind as to agree with a comparative notion of freedom. According to this, that is sometimes called a free effect, the determining physical cause of which lies within the acting thing itself, e.g., that which a projectile performs when it is in free motion, in which case we use the word freedom, because while it is in flight it is not urged by anything external; or as we call the motion of a clock a free motion, because it moves its hands itself, which therefore do not require to be pushed by external force; so although the actions of man are necessarily determined by causes which precede in time, we yet call them free, because these causes are ideas produced by our own faculties, whereby desires are evoked on occasion of circumstances, and hence actions are wrought according to our own pleasure. This is a wretched subterfuge with which some persons still let themselves be put off, and so think they have solved, with a petty word- jugglery, that difficult problem, at the solution of which centuries have laboured in vain, and which can therefore scarcely be found so completely on the surface.”

Kant, *Critique of Practical Reason*, p.66

It is, however, possible that Kant mis-describes Hume’s position: Hume, after all, had said that “voluntary” actions are those that are caused by the *desires* of the agent – and since the movement of the clock and projectile are not caused by desires those items have, we are not in fact forced to count them as free.

A second problem for Hume is the suggestion that he is wrong to concentrate on free action – being able to act according to our desires and choices – when in fact the real problem is free will or freedom of choice. What good is it to show that our actions are free or “voluntary”, when the desires and decisions that bring
them about are fixed (determined) in advance by factors beyond our control? The incompatibilist had arguments which seem to show that our choices and decisions – our “will” – are not under our own control, and hence not free. What Hume needs to do to respond effectively to such arguments is to explain how we can have Free Will, not voluntary action. This point is sometimes put by saying that Hume confuses or conflates Free Will with political freedom – the freedom to do as we like without interference from external sources of control such as a repressive society or a totalitarian state. Political freedom is the freedom to act as we please; but Free Will, they say, concerns our ability to make different choices and have different desires from the ones we actually do have – something which seems to be ruled out by determinism.

Exercise: “Even if determinism is true, I can still do what I want.” Explain in your own words why this statement is true.

Finally, Hume has been criticized on the grounds that his account of the distinction between voluntary and involuntary action is not sophisticated enough, and is vulnerable to counter-examples. Hume suggests that an action should be counted as voluntary (and so “free”) if it is caused by the desires and choices of the agent (in Hume’s phrase, if the agent acts “according to the determinations of his will”); however, two kinds of counter-example threaten. One is where someone is subject to, and acts upon, desires which are compulsive and out of control – for example, in drug addiction or kleptomania. According to Hume’s definition, such people act voluntarily and freely, even though we might want to say that they are not fully in control of their actions in the way a truly free agent would be.

The second kind of counter-example is where my desire to do something results in my performing another action which I had not intended. For example, my desire to attract your attention might cause me to wave at you. But I am at an auction, and my wave also commits me to buy a very expensive painting. Obviously, bidding for this painting was an involuntary action – and action I did not intend to perform – but it was one that was caused by my desires, and so ought to count as “voluntary” on Hume’s definition of what it is for an action to be “voluntary”.

“Guidance control” as a criterion of freedom.
Do you remember the example of the thermostat (used by Simon Blackburn both in the quote in Week 4, and in the edition of “In our Time”)? The thermostat responds to its input (the temperature sensor in the room), and uses this information to control the behaviour of the house’s heating system. So the thermostat in a sense “controls” the heating system, even though its behaviour is in turn controlled by other things (its input from the thermostat and its programming), and even though there is no possible alternative way it could behave in the circumstances. Similarly, we might think that a person has “control” over her actions and acts freely, so long as her decision-making faculty responds to its inputs in an appropriate way. In particular, we are interested in how she responds to reasons. It doesn’t matter whether there are genuine alternative possibilities for action under the circumstances; what matters is that her actions respond to reasons in an appropriate or rational way. This kind of control over one’s actions is called “guidance control” by the contemporary philosopher John Martin Fischer – it is the control we have in virtue of acting as we do in response to reasons. Crucially, it is possible to have this kind of control even if how we respond to reasons is entirely determined by factors outside our control such as the laws of nature acting on neurons in our brain.

A wonderfully colourful example of “reason-responsiveness” is given in the Stanford Encyclopedia’s article on “compatibilism”:
“The most natural way to understand a reasons-responsive theory is in terms of an agent’s responsiveness to reasons. To illustrate, suppose that Frank Zappa plays the banjo of his own free will. According to a reasons-responsive theory, his playing the banjo freely at that time requires that if, in at least some hypothetical cases, he had reason not to, then he would refrain from playing the banjo. For instance, if Jimi Hendrix were to have stepped into Frank’s recording studio and asked Frank to play his electric guitar, Frank would have wanted to make Jimi happy and thus would have gladly put his banjo aside and picked up his electric guitar.”

In the example, it doesn’t seem to matter that it was determined that, faced with certain reasons, Frank Zappa was guaranteed to have performed certain actions (and couldn’t have done anything else in those precise circumstances); what matters is that Frank is “responsive to reasons” in the sense that he would have acted differently if he was in a different situation and had good enough reasons to act differently. Conversely, anyone who does not respond to reasons in the way we would expect (e.g. if they are compulsive, neurotic, or simply insane) is not counted as having Free Will according to this account, since such people do not respond to reasons in the right way.

Again, this account can be challenged: first, it might be argued that guidance control is not enough (not “sufficient”) for Free Will, since the kind of control over our actions that we want involves the ability to select between different alternatives in the precise circumstances; but guidance control does not give us that – we are free only insofar as we could have acted differently if we’d been in a different situation where we’d had different reasons. Think of Frank Zappa again: surely for him to have played the banjo of his own Free Will we don’t simply need (i) if he’d been in a different situation, with different reasons, he’d have done something else (guidance control); we also need the possibility that in the very situation he was in, faced with exactly the same reasons, he could have done something else. This kind of control is called regulative control, and unlike guidance control it seems entirely incompatible with determinism, since determinism says that, in those precise circumstances, there was only one course of action open to Frank, namely playing the banjo.

We can restate that complicated criticism in more straightforward language: philosophers who think that guidance control is a criterion or mark of Free Will accept that we can be free even it is already fixed (determined) that we will respond to our reasons in a certain way; but it seems that, to be truly free, we must also be able to choose between different ways of responding to the reasons that we are faced with – and that seems to require that there are genuine alternative possibilities for action in the exact circumstances we are in, something that the determinist denies.

A second criticism is that this account rules out the possibility of someone freely choosing to ignore the reasons he is presented with. Someone who is extremely stubborn or bigoted might have plenty of good reasons to change his behaviour – but he chooses not to. We don’t want our definition of Free Will to force us to conclude that such people are not acting freely, simply because they are not responding to reasons; in fact, we count their decision to ignore the reasons they have as an expression of their own free choice. This seems to be a powerful counterexample to the view that responsiveness to reasons is the mark of Free Will.

Second-order desires
We can distinguish between first-order desires (normal desires) and second-order desires – desires about my desires. For example, someone might be in love and so want to spend some time with their beloved; but she might also recognize that the relationship was no good for her, and so want not to be in love – she
might want not to have the desires that she has. The contemporary philosopher Harry Frankfurt has suggested that this distinction shows us how to produce a compatibilist definition of freedom: we are free when our second-order desires are capable of controlling our first-order desires; when we want what we want to want. To understand this better, consider an example of someone who is not free in this sense, because there is a conflict between his second-order and first-order desires:

Consider a “drug addict [who] hates his addiction and always struggles desperately, although to no avail, against its thrust. He tries everything that he thinks might enable him to overcome his desires for the drug. But these desires are too powerful to withstand, and invariably, in the end, they conquer him. He is an unwilling addict, hopelessly violated by his own desires.”

Frankfurt, Freedom of the Will and the Concept of a Person, p.687-8

Frankfurt argues that this drug addict is not free, because although he wants to take the drug (first-order desire), he also wants not to want to take the drug (second-order desire) – but the second-order desire is not powerful enough to control his first-order desires and enable him to stop taking the drug. However, there could be such a thing as an addict who does act freely: such a person would be someone who both wanted to take the drug (first-order desire) and wants to be a drug addict – i.e. they want to want to take the drug (second-order desire).

Exercise: come up with your own examples of cases (i) where someone’s second-order desires do not control their first-order desires (not free); and (ii) where second-order desires do control first-order desires (free).

Freedom, according to this account, consists in a certain kind of relationship between your second-order and first-order desires: if your first-order desires “line up” with your second-order desires – if you have the first-orders desires you want to have – then you have Free Will. Here’s how Frankfurt puts it:

“the statement that a person enjoys freedom of the will means... that he is free to want what he wants to want. More precisely, it means that he is free to will what he wants to will, or to have the will he wants. Just as the question about the freedom of the agent’s action has to do with whether it is the action he wants to perform, so the question about the freedom of his will has to do with whether it is the will he wants to have.”

Frankfurt, op. cit., p.690

You should notice that Frankfurt’s account of Free Will is entirely compatible with determinism: what matters is only whether our second-order desires control our first-order desires; we can be “free” in that sense even if everything about the desires I actually have and the relationship between them was determined by the laws of nature and the state of the universe long ago. According to this account, we can have Free Will even if the fact that we do have Free Will was determined since the beginning of the universe.

One way of criticizing this definition of Free Will is to say that it forces us to treat people whose desires are compulsive, addictive or neurotic as not having Free Will. Such people not only do things that they would (all things considered) rather not do; they also have desires that they would rather not have. So they have a second-order desire (the desire not to have their compulsive desires) which is not powerful enough to control their first-order compulsive desires. Since their second-order desires are not in control of their first-order desires, they are not free. But we might be tempted to think of addicts and neurotics as acting freely (if irrationally). If you think that such people should be counted as acting freely, then you might prefer Hume’s account, according to which any action “in accordance with” our desires counts as a free action, no matter how compulsive or irrational those desires are.
Moreover, the definition of Free Will in terms of second-order desires seems to suggest that all of us have limited freedom, since all of us have at least some second-order desires which are not powerful enough to change our first-order desires. Speaking personally, I would like to have a stronger desire to spend time reading 19\textsuperscript{th} century fiction. But that second-order desire is not powerful enough to change the fact that, given the opportunity to read \textit{Bleak House}, I would rather do something else instead (e.g. watch the TV adaptation). So my second-order desire (wanting to want to read more 19\textsuperscript{th} century fiction) is not in control of my first-order desires. But I don’t want to conclude, from this fact, that my choice of reading material is not a free one, or that I do not have Free Will.

A final problem for Frankfurt’s account is that, on his definition of Free Will, we are free regardless of where our second-order desires come from – what matters is only whether those desires are strong enough to control our first-order desires. But suppose you had a powerful second-order desire implanted in your brain by a mad scientist or hypnotist (e.g. the desire to become a drug addict). Or suppose that there could be a drug so addictive that it creates not only the first-order desire to take the drug, but also a second-order desire to want to be addicted to it. In these circumstances Frankfurt seems committed to saying that we would be free – so long as those second-order desires were strong enough – but that doesn’t seem right: surely if my second-order desires are simply the result of a drug I have taken, then I am not in control of what I want and do not count as having Free Will.
7 - Determinism without Responsibility

Aims: to understand why determinism might be incompatible with moral responsibility; to understand how praise, blame, and punishment can meaningfully be employed if determinism is true. To understand how it might be possible to adopt a hard determinist approach to moral responsibility.

One of the primary reasons for taking an interest in determinism is its connection to questions of moral responsibility and the justification of punishment. You might think: if determinism is true, and in the circumstances I couldn’t have done anything else, then I am not to blame for my actions, and any punishment you might see fit to give me is unjustified. Conversely, if someone does you a favour, you might think that, because determinism is true, they couldn’t have failed to do you that favour; so they have done nothing praiseworthy, and any thanks or gratitude you might feel towards them is unjustified. At least on the face of it, the truth of determinism threatens to upset our whole approach to each other and to ourselves: it is not only praise, gratitude and blame that determinism threatens, but also regret (about things we should have done but didn’t) and guilt (about things we shouldn’t have done but did). If there was only one way I could have acted under the circumstances, why should I trouble myself with thoughts about how things could have been different: things couldn’t have been different unless my circumstances were different, and I’m not in control of those circumstances.

Obviously, there is going to be a close connection between our views on Free Will, and our views about human responsibility. Many philosophers use our judgements about moral responsibility as evidence to try to decide what Free Will is, and whether we have it; some philosophers would even go so far as to say that Free Will just is whatever property human beings have that makes us morally responsible – thus they say that there is a conceptual connection between the idea of Free Will and the idea of moral responsibility. Because of this, it is often thought acceptable to use arguments about whether we are “free” to try to decide a question about whether we have moral responsibility. Other philosophers would say that the expression “Free Will” is not worth arguing about, and we should debate only the connection between determinism and moral responsibility, without worrying about whether we can call ourselves “free” or not.

On the whole, the debate about determinism and responsibility divides up into factions you have already met: compatibilists generally argue that determinism and moral responsibility are compatible – that we can be morally responsible even if determinism is true – while incompatibilists say that determinism and moral responsibility are incompatible – so we need to decide whether we want to keep moral responsibility and reject determinism (libertarians) or keep determinism and reject moral responsibility (hard determinists). However, this neat picture can become a little more complex, since some hard determinists (who claim that we have no Free Will and are not ultimately morally responsible) nevertheless argue that we might have good reason to go on using our practices of praise and blame, even if people are not free to act differently and strictly speaking are not responsible for their actions.

Arguing for incompatibility: “ought” implies “can”

Suppose two people – Bill and Fred – are walking on a beach and they see a drowning child far out to sea. Bill says to Fred “You ought to jump in and rescue him”; Fred replies “but I can’t swim!” Most people would agree that Fred is not morally responsible for his failure to save the drowning child by swimming out to him, since there was no way he could have saved him – he just couldn’t have swum out far enough to save the child. However, if there were some other way of saving the child that Fred could have used but didn’t (for example using his mobile phone to call the coastguard) then this is something that we do hold him morally responsible for.
Examples such as these are used to support the principle that you can be held morally responsible for something only if you could have avoided doing it – or in short, that “ought” implies “can”: saying that Fred ought to have saved the drowning child implies that Fred could have saved the drowning child, and if Fred genuinely couldn’t have saved the drowning child, then we no longer claim that Fred ought to have saved the drowning child.

How does this connect to determinism? The thought is simple: according to the determinist, everything that happens is determined (fixed in advance) by the previous state of the universe plus the laws of nature; so, given the circumstances now, there is only one way things could turn out next. This means that, in the circumstances, Fred couldn’t have done anything other than he actually did; in which case (since “ought” implies “can”), we can’t blame Fred for not acting differently – since strictly speaking, in the circumstances he couldn’t have acted differently. Suppose Fred is a strong swimmer who knows it will be easy for him to save the drowning child, but he still lets the child drown because he is a lazy, unpleasant man. Even then we can’t hold Fred responsible for his failure to save the child, since – given the truth of determinism – Fred literally could not have behaved any differently from how he actually behaved. We might wish that Fred had behaved differently and saved the child; however, we cannot blame him for not saving the child, because we can’t criticize someone for failing to do something it was impossible for them to do in the circumstances. So “ought” implies “can” is a principle appealed to by incompatibilists who believe that determinism and moral responsibility cannot coexist. In particular it is often appealed to by hard determinists who argue that, because determinism is true, no-one is ultimately responsible for anything.

However, there are ways to respond to this argument. One is provided by Daniel Dennett, who points out that our usual standards of praise and blame are not abandoned when an agent “could not do otherwise”: according to Dennett, the “ought” implies “can” principle should be rejected, since there are clear counter-examples to it. Most memorably, there is the example of Martin Luther’s statement taking full responsibility for his actions when challenged about his anti-clerical writings at the “Diet of Worms” in 1521: “Here I stand – I can do no other.” The point is that Luther’s feeling of conscience over the state of the church made it literally impossible for him to act differently from how he actually behaved. We might wish that Fred had behaved differently and saved the child; however, we cannot blame him for not saving the child, because we can’t criticize someone for failing to do something it was impossible for them to do in the circumstances. So “ought” implies “can” is a principle appealed to by incompatibilists who believe that determinism and moral responsibility cannot coexist. In particular it is often appealed to by hard determinists who argue that, because determinism is true, no-one is ultimately responsible for anything.

Dennett, I Could not have Done Otherwise – So What? (1984) p.556

(Dennett’s argument has some parallels with the “natural language philosophy” popular at Oxford in the 1940s and 50s: he uses facts about everyday speech and how we normally use words to try to show that a philosophical principle is mistaken. More on that next week.)

[Further reading: Dennett, I Could not have Done Otherwise – So What? p.555-7]

A second problem with the “ought” implies “can” principle is that its use in arguments about responsibility
can be self-defeating. Suppose that the incompatibilist wants to use this principle in an argument to show us how we ought to think about moral responsibility. If it is true that “ought” implies “can” then it can be true that we ought to think about moral responsibility differently only if we can think about moral responsibility differently. But if determinism is true, then our views about moral responsibility are fixed in advance: given the laws of nature and the original state of the universe, there is only one way I could think about moral responsibility at now or at any point in the future. So – if we believe that determinism is true and that “ought” implies “can” – it seems paradoxical and self-defeating to say that we ought to change our views on moral responsibility. Strictly speaking, no-one ought to change her views, since strictly speaking no-one can change her views.

Question: This is only a problem for only one faction of incompatibilists: which one is it – libertarians or hard determinists? Why?

Alternative justifications for praise and blame
Suppose you agree with the “ought” implies “can” principle, and so believe that determinism is incompatible with moral responsibility: that, because we could not have done otherwise in the circumstances, we do not bear ultimate moral responsibility for our actions. Nevertheless, it would be a mistake to conclude too hastily that our everyday practices of praising and blaming people are unjustified and should be avoided where possible. Even if determinism is true, it is still the case that what we say and do has an influence on how people behave: we are still capable of changing the behaviour of those around us for the better, even if what we do to try to change their behaviour is fixed in advance by factors outside our control. This enables determinists to give an account of how praise, blame, and punishment can meaningfully be employed if determinism is true. The reason we are justified in praising and blaming people (and punishing them, sometimes) can be that it enables us to influence their future behaviour for the better. Determinists believe that everything a person does is fixed in advance; but that is precisely because they think of people as complex systems that are subject to causal influences from outside. So it is possible for an outside input (e.g. my praising you for what you have done) to influence the complex reasoning system embedded in your brain so that you are more likely to carry out such praiseworthy acts in future. In fact, the act of praising someone (if it is witnessed by others) can have a positive effect on more people than merely the person who is praised.

Scientists and mathematicians who work in Artificial Intelligence call this kind of system a feedback loop – a system that passes information back into the “decision-making” centre of the programme and so influences the machine’s behaviour in the future. Here’s an example of how that might work with a human:

Lola (aged two) steals a packet of sweets from the newsagents. Her parents tell her that this was “wrong” and that she is “naughty”. They are also very cross with her; all in all she has an unpleasant time for the rest of the day. This experience is fed back into the decision-making parts of her neural network, and adjusts the relative priorities her decision-making faculty assigns to (i) not stealing and (ii) having sweets. From now on, where there is a choice
between having sweets and not stealing, Lola will choose not to steal.

That example used a toddler for simplicity’s sake; but we could easily imagine a similar justification of blame or praise without moral responsibility even when the person on the receiving end is a mature adult. This kind of justification is a **pragmatic** justification: it justifies our behaviour in praising or blaming someone on the basis of whether that behaviour “gets results”, regardless of whether the people involved really do deserve that praise or blame.

**Exercise:** explain how the behaviour towards individuals in these examples could be justified pragmatically:

- Seamus Heaney is honoured by being given the Nobel Prize for Literature.
- Charlie Gilmour is imprisoned for breaking windows and throwing a bin at Prince Charles’ car.
- Captain Alfred Shout is awarded the VC posthumously in 1915 for hand-to-hand combat in the trenches in Gallipoli, Turkey.
- Everyone applauds Aisling and Emma’s performance at PopFest.
- Lilibet feels guilty about not handing her homework in on time.

The problem with this strategy for saving our normal ways of praising and blaming people is that it does not deal with the accusation that punishing people for things they are not responsible for is unfair. After all, if someone is not (really) responsible for her actions because she couldn’t have done anything else in the circumstances, then she does not deserve to be punished. Suppose we all agree that punishing her is useful for society, and will “do her good” in the sense of making her better behaved in the future; nevertheless it isn’t fair on her, since she wasn’t free to act any differently and it wasn’t her fault that she acted as she did! Similarly, praising people for their achievements also begins to seem unfair on the people who don’t get praised – never mind that praising people will encourage others to follow their example; it is still the case that the people who get praised can’t justifiably take the credit, because they could not have failed to act as they did.

**The consolations of hard determinism**

The most extreme form of hard determinism is one that simply says: we are not free, and we are not ultimately morally responsible for what we do and the sort of person we are, so get used to it; the entire project of praising and criticizing other people from a moral point of view is a mistake. (This kind of position is known as an **error-theory** about morality.) How, then, should we live? The contemporary hard determinist Ted **Honderich** has made some suggestions about what a hard determinist worldview would be like. Obviously, he says, we have to give up “life-hopes” which are linked to the idea that the future might somehow be open and not yet determined, or based on the idea that we might be the ultimate **originators** of our actions:

“All our life-hopes involve thoughts to the effect that we somehow initiate our future actions. Some involve not only beliefs as to voluntariness or willingness but also an idea, or what is more an image, of our originating our future actions. To think of life-hopes of this kind, and their manifest inconsistency with determinism, and to accept the likely truth of determinism, is to fall into dismay. We are deprived of the hopes.”

Honderich, *The Consequences of Determinism*, p.169
Moreover, we should not accept any moral or legal framework according to which moral responsibility is irrelevant to punishment:

“We should have resist or anyway think hard about the increas of strict-liability legislation and interpretation, and have no truck with any supposedly enlightened proposal that an offender’s degree of responsibility for an act should be irrelevant to whether he has broken the law and to what happens to him.”

Honderich, *On Determinism and Freedom*, p.67

Strict liability is the legal term for a situation in which someone is liable for his actions regardless of whether he is morally responsible for them; Honderich’s point here is that we should not go on holding people criminally liable for their actions even though we have decided that they do not have ultimate moral responsibility for those actions.

Nevertheless, he claims, even the hard determinist can find ways to avoid this despair: we can meaningfully hope that our actions will be consistent with our character, and that our desires will be fulfilled, even if we are not ultimately responsible for that character and those desires. We can hope:

“that we will act not from reluctant desires and intentions, but from embraced desires and intentions, that we will act in enabling circumstances rather than frustrating ones. These circumstances have to do with at least the way of my world, the absence of self-frustration, independence of others, and absence of bodily constraint. ... These life-hopes are not at all significantly threatened by determinism.”

Honderich, *The Consequences of Determinism*, p.169

Moreover, there are some ways in which embracing a hard determinist world-view may make our lives more rather than less comfortable: it is true that we will no longer take ultimate credit for our successes; but we will no longer feel any sense of guilt for our failures. We cannot congratulate ourselves for reaching the top of our profession; but the majority of people, who do not reach the top of their professions, need not feel bad about the fact that they did not succeed as much as they could (for they could not have succeeded any more). When a teacher writes “could try harder”, the hard determinist pupil can take comfort from the fact that this is not true – she could not have tried harder in the circumstances.

Similarly, we might find it easier to deal with other people. We can no longer feel gratitude to our friends; but equally we need no longer feel any resentment towards those who have wronged us: it was fixed from the start of the universe that they were going to wrong us, and they are not responsible for that fact. Taking vengeance on them for their actions would be unjustified – but people who are always seeking retribution for wrongs done them are often bitter and miserable. Better, surely, to use hard determinism as a way of helping us “forgive and forget”.

[further reading: Galen Strawson, “On Free Will”]
8 – Determinism with Responsibility

Aim: to understand the “compatibilist” view that moral responsibility is still possible in a deterministic universe.

Moral responsibility and “ownership” of my actions

You’ll remember that Hume defined “voluntary” action in terms of the cause from which it issued: an action is voluntary, and hence free, if it is “in accordance with the determinations of the will” – i.e. if the action was caused by my desires and choices in an appropriate way. Unsurprisingly, Hume offers a related test for moral responsibility: actions can be praised or blamed (can be “objects of our moral sentiment”) if they proceed from our “internal character, passions, and affectations”; but actions cannot be praised or blamed if they are not caused by something internal to me, but rather “are derived altogether from external violence”:

“actions are objects of our moral sentiment, so far only as they are indications of the internal character, passions, and affections; it is impossible that they can give rise either to praise or blame, where they proceed not from these principles, but are derived altogether from external violence.” (Hume, Enquiry 8.ii)

For example, I am not to be blamed (or praised) for hitting Fred if Bill grabbed my arm and swung it towards Fred, as then my action would proceed from “external violence”; however, if the cause of my action was my deep and abiding hatred of Fred, then my action would proceed from my “internal character” and so it would be something I can be praised and blamed for.

You should be aware, however, that Hume makes exceptions: some actions which are caused by my own desires, although they are “voluntary” and freely chosen, are not subject to praise and blame, because they are out of character, and do not reveal the underlying “character and disposition” of the person:

“Actions are by their very nature temporary and perishing; and where they proceed not from some cause in the character and disposition of the person, who perform’d them, they infix not themselves upon him, and can neither redound to his honour, if good, nor infamy, if evil. The action itself may be blameable ... But the person is not responsible for it; and as it proceeded from nothing in him, that is durable or constant, and leaves nothing of that nature behind it, 'tis impossible he can, upon its account, become the object of punishment or vengeance.” (Hume, Treatise 2.3.2)

Hume’s reason for saying this is that he believes that, strictly speaking, we apply the “sentiments” of praise and blame not to the actions themselves, but to the character that produced them. We only criticize actions in so far as they are “signs or indications of certain principles in the mind or temper”; it is the mind that produced them that is the proper object of our feelings of praise or blame:

“'Tis evident, that when we praise any actions, we regard only the motives that produced them, and consider the actions as signs or indications of certain principles in the mind and temper. The external performance has no merit. We must look within to find the moral quality. This we cannot do directly; and therefore fix our attention on actions, as on external signs. But these actions are still considered as signs; and the ultimate object of our praise and approbation is the motive, that produc’d them.” (Hume, Treatise, 3.2.1)

So Hume makes two important suggestions about responsibility: first, that an action for which we are morally responsible must be the result of our internal character and not “external violence”; and second, that we are responsible for actions only if they are a good indication of our “mind and temper” and not out-of-character. We might sum this up by saying that Hume is putting forward a view about what it means to have ownership of our actions: we have ownership of those actions which reveal something about our
internal character; these are the actions for which we are rightly praised or blamed.

However, this approach faces a problem (which Hume himself recognized): how can we claim ownership of our character and actions if these were fixed long before we were born, as the determinist claims? If your character is nothing more than the result of our genetic makeup and experiences in life, how can it be fair for me to praise, blame, or punish you for who you are, since the sort of person you are was entirely determined by factors beyond your control? This is the kind of argument which has led “hard determinist” philosophers such as Galen Strawson to say that we should stop holding people morally responsible for their actions, no matter how terrible those actions are – as he memorably puts it, “even Hitler is off the hook”.

Hume’s own answer to this problem is naturalistic – he is more interested in explaining the “sentiments” of praise and blame, and how they arise from human nature, then in trying to justify our behaviour in praising and blaming people from some neutral, objective point of view. For him, it seems, it is good enough that human nature guarantees that we will feel such “sentiments” in certain situations – we do not need to worry too much about whether this behaviour is fair or not, since it’s going to happen whatever philosophical arguments we bring up:

“The mind of man is so formed by nature that, upon the appearance of certain characters, dispositions, and actions, it immediately feels the sentiment of approbation or blame; nor are there any emotions more essential to its frame and constitution.” (Hume, Enquiry 8.ii)

Appeal to natural language; determinism and the law

In the 1940s and 1950s a number of philosophers working at the University of Oxford developed an approach which came to be known as “natural language philosophy”, according to which philosophical problems are fundamentally problems about the way we use words, and can be solved simply by paying closer attention to how we use words. Such philosophers tended to think that the business of philosophy was to understand our natural ways of talking, rather than to criticize them. You might be tempted to adopt this approach to the problem of determinism and responsibility: then the important thing to do will be simply to clarify under what circumstances we hold people responsible for their actions. Usually such appeals to natural language are used to justify a form of compatibilism.

You have already met one important point, made by Dennett: in our ordinary use of language we do not always worry about whether an agent “could have done otherwise” before assigning moral responsibility for his actions. Martin Luther took full responsibility for his actions, even though he felt that in the circumstances his conscience allowed him to do nothing else. Instead, the natural language philosopher will point out that in everyday speech we have different tests for whether a person should be held responsible for his actions, for example:

- Did he understand the likely outcome of his action?
- Did he have an intention to do what he did?
- Was he under emotional duress?
- Did he understand the situation he was in?
- Had he reached intellectual and emotional maturity?

Someone could satisfy all of these tests for responsibility even in a deterministic world; so, the natural language philosopher claims, our everyday use of language shows us that people can be responsible for their actions even if determinism is true.
Another important point about ordinary language, made by Oxford philosopher J.L. Austin, is that in everyday life (and especially in criminal courts) moral responsibility is not an all-or-nothing affair, since in murder trials we allow a defence of diminished responsibility in which the guilty party is held responsible for the crime of manslaughter, but not for murder.

Exercise: the law on diminished responsibility changed in 2009. What significant differences can you find, and can you explain them?

Persons suffering from Diminished Responsibility (England and Wales)

Homicide Act 1957, s.2:

Where a person kills or is party to a killing of another, he shall not be convicted of murder if he was suffering from such abnormality of mind (whether arising from a condition of arrested or retarded development of mind or any inherent causes or induced by disease or injury) as substantially impaired his mental responsibility for his acts and omissions in doing or being a party to the killing.

Coroners and Justice Act 2009, s.52:

(1) A person (“D”) who kills or is a party to the killing of another is not to be convicted of murder if D was suffering from an abnormality of mental functioning which—

(a) arose from a recognised medical condition,
(b) substantially impaired D’s ability to do one or more of the things mentioned in subsection (1A), and
(c) provides an explanation for D’s acts and omissions in doing or being a party to the killing.

(1A) Those things are—

(a) to understand the nature of D’s conduct;
(b) to form a rational judgment;
(c) to exercise self-control.

Nevertheless, there are problems with appealing to ordinary language to settle the question about responsibility in favour of compatibilism. It is true that, in our ordinary use of language, we hold people responsible for their actions without worrying about whether those actions were fixed in advance, as the determinist claims. But it would be too quick to conclude from this that the truth of determinism makes no difference to the question whether we are morally responsible. One problem is this: describing how we do use language does not help us understand how we should use language. Even the hard determinist agrees that people talk as though the truth of determinism is irrelevant to moral responsibility; but, he says, that just shows that people haven’t thought through the issue! We can’t settle a philosophical debate simply by appealing to “common sense” or “what most people think”, since people can easily be mistaken. (It used to be the case that most people thought that the Earth was flat.)

A second problem is that appealing to natural language might show us that our ordinary moral judgements are, after all, incompatible with determinism. A very famous example of this is known as the “Twinkie defence”. In San Francisco in 1979 Dan White was convicted of voluntary manslaughter rather than murder after offering a defence of “diminished capacity” on the basis of depression. In the trial, White’s change of diet to sugary foods (including Twinkies) had been cited as a symptom of this depression; however, one of the psychiatrists called as expert witnesses also mentioned that the change of diet might have made his depression worse. Media reports at the time exaggerated the significance of the suggestion that the high
sugar intake caused the murder, and the verdict provoked riots; however, the fact remains that Dan White escaped the death penalty because of depression – a chemical disorder of the brain. The idea that the chemical state of our brain can excuse us from responsibility for our actions is, in fact, commonplace: people often say things like “that’s the drink talking” to excuse unfortunate comments, and may even say “it wasn’t me; it was the drink”. Similarly we might talk about actions being caused by a “caffeine buzz”, or a “sugar rush”. The fact that some of us allow these claims to go unchallenged suggest that we do not always hold people responsible for actions which can be traced back to the chemical state of the brain. But the determinist suggests that every action can in principle be traced back to some prior cause which necessitates it – perhaps even to a specific chemical state of the brain. So perhaps, ultimately, common sense will force us to conclude that, because every action is necessitated by brain chemistry, no action is such that we are responsible for it.

**The intentional stance: causes and reasons**

An alternative compatibilist response is provided by the language of different “stances” you met in week 1. Holding someone responsible for her actions is part of taking the **intentional stance** towards her – i.e. treating her as a creature which forms intentions and acts on them, who responds to reasons, and whose behaviour can be explained in terms of beliefs and desires. This is contrasted with the **mechanical stance**, which involves explaining her behaviour in terms of the physical causes which brought it about – for example, low serotonin, high bloodsugar, malfunctioning neurons in one part of her brain. Crucially, however, we choose to adopt different stances to the same thing. Moreover, the choice of stance is justified **pragmatically** – we choose the stance which gives the best results when predicting behaviour. So we are justified in praising and blaming someone so long as we are justified in taking the intentional stance towards her; at the point at which the intentional stance is no longer useful for predicting her behaviour we fall back to the mechanical stance. This would be the same switch as is made when someone switches from treating a mental illness psychologically (by investigating the reasons for the patient thinking as she does) to treating it psychiatrically (by medication designed to control brain chemistry).

The pragmatic nature of the choice of “stances” explains why we sometimes apply the “intentional stance” to inanimate objects. You will probably have greater success in playing computer games if you try to work out what the game “wants” or “expects” you to do – in other words, if you approach the game using the intentional stance. Moreover, if it is possible to treat even computer games – whose behaviour is determined by their programming – using the intentional stance, it seems that the fact that our behaviour is determined in advance is no obstacle to treating each other using the intentional stance; and if we are justified in using the intentional stance when dealing with each other, we are justified in praising and blaming each other.

**Question:** Is the mere fact that the intentional stance is useful in predicting human behaviour a good enough reason to go on praising and blaming people for their actions?

**Strawson and the “reactive attitudes”**

The 20th century philosopher P.F. Strawson (father of Galen Strawson, who you have met a few times already) proposed a novel approach to the problem of determinism and responsibility. His strategy was to argue that it is a **practical impossibility** for us to stop holding other people morally responsible, so long as we continue thinking of them as people and fellow members of society to whom we can adopt the “normal interpersonal attitudes”. To recognize someone as a person just is to treat him as a fitting object of praise, blame, gratitude and resentment; giving up the right to praise and blame someone would require us to
treat him as some kind of machine to be manipulated as our purposes dictate. Moreover, we could be persuaded us to take this inhuman “objective attitude” to someone else – and thus to stop treating them as a person – only if the person had some mental illness or other problem that made him irrational or incapable of participating in normal human interaction. Crucially, Strawson is convinced that our “theoretical” belief in the truth of determinism cannot stop us treating other people as capable of normal interpersonal interaction; and since treating people in this “normal” way requires a belief in moral responsibility, he concludes that the truth of determinism is no threat to our usual practices of holding people responsible for their actions. Thus Strawson’s view is a kind of compatibilism: he argues that a belief in determinism does not – in fact, cannot – make us stop holding other people responsible for their actions.

Exercise: read this extract from Strawson’s paper and see how much of it you can understand.

Selection from Strawson, *Freedom and Resentment* (1962):

“What I want to contrast is the attitude (or range of attitudes) of involvement or participation in a human relationship, on the one hand, and what might be called the objective attitude (or range of attitudes) to another human being, on the other. Even in the same situation, I must add, they are not altogether exclusive of each other; but they are, profoundly, opposed to each other.

To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided ... The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. If your attitude towards someone is wholly objective, then though you may fight him, you cannot quarrel with him, and though you may talk to him, even negotiate with him, you cannot reason with him. You can at most pretend to quarrel, or to reason, with him.

Seeing someone, then, as warped or deranged or compulsive in behaviour or peculiarly unfortunate in his formative circumstances—seeing someone so tends, at least to some extent ... tends to promote, at least in the civilized, objective attitudes. But there is something curious to add to this. The objective attitude is not only something we naturally tend to fall into in cases like these, where participant attitudes are partially or wholly inhibited by abnormalities or by immaturity, It is also something which is available as a resource in other cases too. We look with an objective eye on the compulsive behaviour of the neurotic or the tiresome behaviour of a very young child, thinking in terms of treatment or training. But we can sometimes look with something like the same eye on the behaviour of the normal and the mature. We have this resource and can sometimes use it; as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity. Being human, we cannot, in the normal case, do this for long, or altogether. If the strains of involvement, say, continue to be too great, then we have to do something else -- like severing a relationship. ...
Our question reduces to this: could, or should, the acceptance of the determinist thesis lead us always to look on everyone exclusively in this way? ... I am strongly inclined to think that it is, for us as we are, practically inconceivable. The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of ... attitudes and feelings that is in question.

This, then, is a part of the reply to our question. A sustained objectivity of inter-personal attitude, and the human isolation which that would entail, does not seem to be something of which human beings would be capable, even if some general truth were a theoretical ground for it. But this is not all. There is a further point, implicit in the foregoing, which must be made explicit. Exceptionally, I have said, we can have direct dealings with human beings without any degree of personal involvement, treating them simply as creatures to be handled in our own interest, or our side’s, or society’s—or even theirs. In the extreme case of the mentally deranged, it is easy to see the connection between the possibility of a wholly objective attitude and the impossibility of what we understand by ordinary interpersonal relationships. Given this latter impossibility, no other civilized attitude is available than that of viewing the deranged person simply as something to be understood and controlled in the most desirable fashion. To view him as outside the reach of personal relationships is already, for the civilized, to view him in this way. For reasons of policy or self-protection we may have occasion, perhaps temporary, to adopt a fundamentally similar attitude to a ‘normal’ human being; to concentrate, that is, on understanding ‘how he works’, with a view to determining our policy accordingly, or to finding in that very understanding a relief from the strains of involvement. Now it is certainly true that in the case of the abnormal, though not in the case of the normal, our adoption of the objective attitude is a consequence of our viewing the agent as incapacitated in some or all respects for ordinary interpersonal relationships. He is thus incapacitated, perhaps, by the fact that his picture of reality is pure fantasy, that he does not, in a sense, live in the real world at all; or by the fact that his behaviour is, in part, an unrealistic acting out of unconscious purposes; or by the fact that he is an idiot, or a moral idiot.

But there is something else which ... is equally certainly not true. And that is that there is a sense of ‘determined’ such that (1) if determinism is true, all behaviour is determined in this sense, and (2) determinism might be true, ... and (3) our adoption of the objective attitude towards the abnormal is the result of a prior embracing of the belief that the behaviour ... of the human being in question is determined in this sense. Neither in the case of the normal, then, nor in the case of the abnormal is it true that, when we adopt an objective attitude, we do so because we hold such a belief [in determinism].

So my answer has two parts. The first is that we cannot, as we are, seriously envisage ourselves adopting a thoroughgoing objectivity of attitude to others as a result of theoretical conviction of the truth of determinism; and the second is that when we do in fact adopt such an attitude in a particular case, our doing so is not the consequence of a theoretical conviction which might be expressed as ‘Determinism in this case’, but is a consequence of our abandoning, for different reasons in different cases, the ordinary inter-personal attitudes.”