

RESPONSIBILITY, LIABILITY, AND RETRIBUTION

Jaap Hage & Antonia Waltermann

1. Responsibility and Liability

1.1 The Roles of Responsibility and Liability

On his way to bring his 5-year old daughter Susan to school, John is driving his car through the quiet outskirts of his hometown. Suddenly, Susan starts making unexpected noises, so John turns to the back of the car to see what is going on. It is nothing serious – but while John is trying to figure out what is going on, a pedestrian, Geraldine, crosses the street. Because he is not paying attention, John notices Geraldine too late. He cannot brake in time and hits Geraldine with his car. Fortunately, Geraldine survives the accident, but not without injuries which require hospitalisation. Even though he was distracted by his young daughter, John is still responsible for not braking in time, and he is liable for at least part of Geraldine's medical costs. Moreover, John is responsible for violating traffic rules and, as a consequence, he is liable to be fined.

As the short story above illustrates, the notions 'responsibility' and 'liability' play important roles in law. John's responsibility for not braking in time and the ensuing liability for damages illustrate their role in private law, in particular the law of torts. His responsibility for violating the rules of traffic law and the ensuing liability to be fined illustrate a similar role for these notions in criminal law. In both cases, John's liability depends on his responsibility for what he did. This relationship between liability and responsibility – with liability depending on and being a consequence of responsibility – is typical for a retributivist view on liability: an agent is liable for damages or liable to be punished because she deserves it.¹ This central role of desert is characteristic of retributivism.

This retributivist view of liability has been relatively unproblematic for a long time. In recent decades, however, this has begun to change. The nature of responsibility has become a hotly debated issue in moral and legal philosophy (e.g. Hart, 1968; Feinberg, 1970; White, 1985; Fischer and Ravizza, 1993; Kenneth, 2001; Lucy, 2007; Sinnott-Armstrong and Nadel, 2011; Vincent, 2013; Sinnott-Armstrong, 2014). Moreover, the dependence of liability on responsibility has become questionable. One of the reasons why the notion of responsibility has been problematized is because of the increasing insights into the causes of human behaviour that the flourishing of the cognitive sciences has brought about. The debate whether human beings can ever be responsible for their actions is as old as debates about

¹ In this paper, we will apply the convention that an author uses the pronouns that reflect her or his own gender. In the case of this co-authored chapter, this means that male and female pronouns are used interchangeably.

the truth of determinism. It has become topical since the cognitive sciences have provided evidence of how this alleged determination is – literally – embodied (Greene & Cohen, 2004).

The central topics of this chapter are whether responsibility exists and whether law – in particular criminal law – should base liability on responsibility and responsibility on retribution. These questions have received considerable attention in recent decades, and it is not our aim to summarise these extensive debates.² Instead, we want to briefly present the main lines of the debates and then examine whether – and if so, how – the different positions within the debate can be reconciled.

In this examination, a central role is reserved for a social practice we call the ‘practice of agency’ and the tension between two different ways of looking at the world around us, namely the phenomenological and the realist way. Before we can consider a possible reconciliation, however, we first want to introduce and disambiguate the central notions of responsibility and liability and explain why retributivism plays such an important role in the arguments surrounding these notions. Section 1.2. of this chapter will specify what we mean by ‘responsibility’; section 1.3. distinguishes two forms of liability, civil and criminal. Criminal liability can be justified by reference to consequentialism or retributivism, which are introduced in section 1.4. Retributivism makes presuppositions that go hand in hand with free will libertarianism, one of the positions in the debates. We outline this position in section 2, before turning to the arguments against it in section 3. Section 4 describes compatibilism as an attempt to reconcile opposing positions – free will libertarianism, retributivism, and the phenomenological way of looking at the world on the one hand, and hard determinism, no responsibility, and the realist perspective on the other. In section 5, we examine whether compatibilism can succeed in reconciling these positions, and whether such a reconciliation is desirable. We conclude in section 6.

1.2 Responsibility

The term ‘responsibility’ is ambiguous. For instance, Hart (1968, p. 211/2) famously distinguished between four forms of responsibility. Sometimes, the word ‘responsible’ is used as a synonym for ‘liable’. For instance, it is possible to say that John is responsible for the damage suffered by Geraldine. A different use of ‘responsible’ makes the word stand for accountability, as when we say that Harry was not responsible for what he did because he was suffering from severe paranoia at the time of the act. The plural ‘responsibilities’ can be used for the set of duties attached to a particular role, as when we say that the responsibilities of an airplane pilot are manifold.

² Good recent overviews can be found in the work of Kane (2002, 2005) and online (Caruso, 2018; Talbert, 2019; O’Connor and Franklin, 2020).

The notion of responsibility that is at stake in this chapter is closely connected to the notion of agency. We can ascribe an action to an agent by claiming that the agent is responsible for the action. For instance, 'Louise is responsible for the theft' can be used to say that Louise is the one who stole something. Sometimes, agency goes hand in hand with blame or praise for the agent, based on his performing the action. The relation between being considered the agent of some action and deserving praise or blame for the action is so close that we sometimes refuse to call a person responsible for an action if she has an excuse. (From here on, we will ignore praise, because it is less relevant for criminal liability.) For instance, if Alan was blackmailed and under that influence committed theft, we do not blame him for doing so, and this absence of blame is reflected in an unwillingness to call Alan responsible for the theft. Perhaps Alan stole, but the blackmailer is responsible. Moreover, this absence of responsibility will then generally³ be translated into an absence of legal liability: Alan is not criminally liable, because he is not responsible for his action.

In this chapter, we will use the word 'responsible' for the combination of being both the agent and the proper target for blame.

1.3 Liability

As the brief case at the start of this chapter has already illustrated, the law recognises at least two kinds of liability. The first kind is the liability to compensate damage; the second kind is the liability to be punished. As the first kind of liability primarily plays a role in private law, more in particular in tort law and in contract law, we will call this kind 'civil liability'.⁴ In law, punishment is a reaction to crime, and therefore we will call the liability to be punished 'criminal liability'.

Civil liability is the obligation of a legal subject – often a natural person, but it can also be a non-human entity such as a company or a state – to compensate the damage suffered by another legal subject. If the damage was caused by an unlawful action for which the former legal subject can be blamed, we speak of 'fault liability'. In the case described above, John's liability for Geraldine's damage is an example of this. Fault liability can also arise in a contractual setting, where the unlawful behaviour consists in defaulting on one's contractual obligation. For example, if Firm A sold 10,000 widgets to Firm B, but does not deliver in time, Firm A commits contractual default, which is unlawful. If the debtor

³ The relation between responsibility and legal liability depends on the law of a particular jurisdiction. Therefore, it is hard to make statements about this relation that are true without exceptions.

⁴ Liability for damages also plays a role in connection to administrations and states. Some would see this as evidence that liability for damages also plays a role in administrative and in private law. Others see this as a reason why private law also deals with administrations and states. For present purposes, this does not matter.

can be blamed for the non-compliance, this is also a case of fault liability.⁵ Fault liability in civil law is an example of liability dependent on responsibility: an agent was at fault (that is, responsible) and therefore liable.

Fault liability is no longer the only – or even main⁶ – form of civil liability. The requirement that the liable legal subject can be blamed for her action was the standard on the European continent until far into the 20th century. Since then, however, there has been a gradual shift since toward more strict liability, in order to provide the victim of wrongfully caused damage with legal protection and for reasons of economic efficiency.⁷ This shift from fault liability to a stricter form of liability illustrates the increasing role of consequentialist considerations in civil liability.

For example, a legal subject may, under certain circumstances be civilly liable for damage caused by other persons, such as employees, pupils, or children, and in that case we have a stricter form of liability, as this liability may also exist if the liable subject cannot be blamed for causing the damage.⁸ A legal subject may also be liable for damage caused by events that she could not have prevented, and in which no person acted. If, because of a heavy storm, tiles fall off a well-maintained roof, and a passer-by is hurt, the possessor of the building may be liable for the damage, even if he was in no way involved in the accident. However, the possessor would not be responsible for causing the damage in such a case. This is a case of liability without responsibility, exemplifying civil liability based on consequentialist considerations (Van Dam, 2014, Section 1001f.; Hage, 2017; Smits, 2017, Chapter 3). In criminal law, meanwhile, liability is still more strongly related to and dependent on responsibility. Criminal liability is a precondition for punishment. If somebody commits a crime and if this has been established officially, it is possible for an official to impose a sanction on this person. If this possibility exists, we speak of ‘criminal liability’. Although there are a few exceptions, criminal liability typically presupposes that the criminal is considered responsible and blameworthy for the crime (Fletcher 2007, p. 298-339; Blomsma and Roef, 2019b, p. 207). The punishment is then considered to be ‘deserved’ by the criminal. For instance, John violated the traffic rules and therefore deserves having to pay a fine. Firm C intentionally provided the tax service with incorrect information about its profits and in that way dodged taxes. The firm therefore also deserves to be fined. In both cases, the liability and the actual punishment based on it are seen as a way to redress a wrong from the past; punishment is considered as a kind of retribution.

⁵ For historical reasons that go back to Roman law, it is customary to distinguish between contractual and tort liability. However, fault liability and also stricter forms of liability in contract and tort law are so similar that the historical distinction seems outdated.

⁶ This depends on the legal system in question, which makes generalised statements difficult.

⁷ In the common law, this shift has not been as prevalent.

⁸ The liability is called ‘stricter’, rather than ‘strict’, as it may be a condition for vicarious liability that the person who caused the damage could be blamed for what he did.

As the brief discussion above illustrates, criminal and civil liability differ from each other: criminal liability typically requires the responsibility of the criminally liable person, while civil liability is – certainly in the civil law tradition – often a form of stricter liability, which does not require personal responsibility. This difference between criminal and private law can be explained by the different functions liability fulfils in private and criminal law. Civil liability moves damage from the person who suffered it in the first place to the agent that caused it. Whether this is a good idea depends on the interests of all involved parties, including the person who originally suffered the damage. This means that the interests of the person who becomes liable may have to be sacrificed to the interests of the person who suffered the damage. Criminal liability causes additional ‘damage’, not only for the person who will be punished, but also for society, which must enforce the punishment. This damage is added to the ‘damage’ of the crime. Whether this additional damage is justified depends on the interests of society as a whole.

In the rest of this chapter, we will focus on criminal liability, which typically presupposes personal responsibility. Much of the argument, however, is also relevant for fault liability in private law.

1.4. Justification and explanation of liability

In criminal law, there are two main ways to justify criminal liability: consequentialist and retributivist. On a consequentialist approach, the imposition of criminal sanctions is justified by the alleged positive effects of doing so. These effects include the prevention of criminal behaviour through incapacitation or deterrence, reparation of the harm done through crimes, or the rehabilitation of criminals. On a retributivist approach, the imposition of criminal sanctions is justified by the fact that the criminal deserved his punishment because he is responsible for what he did. One factor that plays a role in determining what a criminal deserves is the damage caused by the crime. For instance, genocide deserves a greater punishment than speeding on the highway, and armed robbery deserves a greater punishment than the mere possession of a gun without a permit. Other factors are whether the crime was committed intentionally and the motive for the crime: was it the desire to harm, or merely self-interest?

So far, we have looked at the relationship between responsibility and liability primarily from a normative perspective: what justifies liability? In this connection, retributivism holds that if a person is responsible for some action, this is why she is liable. If we look at the relationship between responsibility and liability from a (cognitive) scientific perspective, however, the question arises whether and to what extent responsibility explains (rather than justifies) why we hold persons liable. More in particular, do the reasons that people adduce to justify their imposition of liability also explain their behaviour?

Carlsmith (2008) found that when people are asked for reasons why they punish, they often mention consequentialist reasons (e.g. deterrence). However, when asked to judge a number of cases that reflected alternatively consequentialist or retributivist reasons for punishing, they imposed punishments in a manner consistent with a retributivist approach to punishment. For instance, they were motivated by the seriousness of the crime, or by the mental state of the agent, more than by the expected (dis)advantages of punishment. This – and much other research⁹ – indicates that retributivist motives play an important role in the decision whether or not to punish. This role is bigger than people think when they reflect on their own reasons for punishing. Retributivist motives, which focus on the responsibility of an agent, may provide a better explanation for the imposition of punishment on this agent than consequentialist ones.

This is an important observation when it comes to debates about the relationship between responsibility and liability. On the one hand, it shows that discussions about responsibility are not merely an academic exercise, but touch upon an important social practice. On the other hand, it shows that scientific findings that question the role of responsibility in attributing liability may have to overcome an important hurdle if they are to influence our social practice, namely the (unconscious) retributivism underlying practices of punishment.

Our natural inclination toward retributivism at first glance seems to be a reason to base responsibility on retribution and liability on responsibility, but this has been called into question. Before we consider the criticism raised against retributivism and retribution as a reason for punishment, we will examine what the retributivist position entails more closely. We do this in section 2, which outlines the position on one side of the debate, namely free will libertarianism. In section 3, we turn to the other side of the debate – hard determinism –, looking at the reasons why retributivism and related positions have been called into question and criticised. Historically, the criticism has been framed terms of determinism and the idea that if determinism is true, there should be no responsibility. As we will see, the argument has lost some of its attraction during the 20th century, when doubts were cast on the universal truth of determinism. However, it has a worthy successor in a related argument about control and agency. Determinism allegedly refutes free will because it allegedly means that an agent has no control over his actions. The argument from control states that irrespective of the truth of determinism, an agent still has no control over his actions and therefore still has no free will. The argument from control is dealt with at the end of Section 3.

Compatibilism seems to offer a solution in the form of a compromise – or rather: compatibility – between the two conflicting positions. In section 4, we first offer a description of compatibilism, before

⁹ For a more extensive overview of relevant research, see section 3 of (Leibovitch, Present Volume).

examining in section 5 whether it really is a solution for the conflict between the libertarian and the hard determinist position.

2. Free Will Libertarianism and Retributivism

In the example at the beginning of this chapter, John has the choice between, on the one hand, paying attention to traffic and complying with the rules of traffic law, and, on the other hand, looking back to see what his daughter is doing, thereby violating the rules. It seems that John could do either, although not both at once, and that he can freely choose which course of action to take. It depends on his will what he will do. This is the idea of free will: an agent has several courses of action open to him, it depends on his will what he will do, and the agent is free to choose. This view of human beings as free agents who are responsible for their actions has been the predominant view, despite (philosophical) challenges in the form of determinism. Running ahead of our argument slightly, determinism denies that agents have more than one option available to them. It is the view that, given the laws of nature, any exhaustive set of states of affairs can only be followed by one particular sequence of new states of affairs (see figure 1). If determinism is true, an agent has, at any given time, only one option available to him. This means that determinism is incompatible with free will as analysed above.

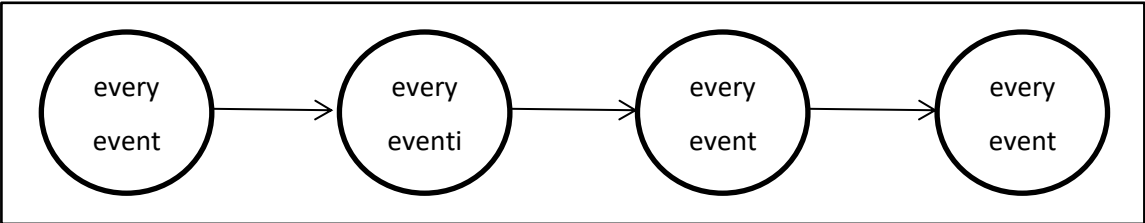


Figure 1: Determinism

Illustrating this, in figure 1, circles represent exhaustive sets of states of affairs – the whole universe – at a particular moment in time, and the arrows represent all the transitions allowed by the set of all laws of nature. This figure represents determinism because every circle, which represents the complete universe at one particular moment, is followed in time by precisely one other circle.¹⁰ This other circle is ‘determined’ by its predecessor in combination with the laws of nature.

Retributivism and responsibility based on retributivism presuppose free will. An agent is free to choose from amongst options and because of this, the agent deserves blame if he chooses wrong. The view

¹⁰ In figure 1, every circle is also preceded by one other circle. That is, as far as determinism is concerned, a coincidence. It is compatible with determinism that different universes at time t determine the same universe at time $t+1$. That would be represented as two different circles pointing to the same follow-up circle.

that defends free will in light of the determinist challenges has become known as ‘free will libertarianism’¹¹ and has recently been defended by Chisholm (1964), O’Connor (1995) and Kane (1996). There are two main strands of free will libertarianism: indeterminism and agent causation. The first denies that determinism is true and typically invokes as reason that quantum physics, which describes entities and events on a micro-level, is in its prevalent interpretation probabilistic and indeterminist (Hodgson, 2002). The traditional counterargument here is that the probabilities and their resulting indeterminacy on the micro-level, if they exist, are cancelled out in macro-phenomena such as neurons and their functioning. That would mean that from a certain level of size on, including the level of actions, determinism would still hold.

2.1. Agent causation

The second strand of free will libertarianism is a modern form of mind/body-dualism, which goes under the name of ‘agent causation’. When we speak of causation, we often do so to ascribe some fact to an action or other event. In common parlance, we typically attribute the status of cause to events and, as a special case, to actions. For instance, the cause of the economic recession in 2020 was the Covid-19 pandemic; the death of Hattie Carroll was caused by William Zanzinger, who hit Hattie with his cane.¹² This last example illustrates that an action can be the cause of an (other) event, but that, at the same time, the agent can also be seen as the cause. This is a typical example of agent causation. Since events, actions and agents are, logically speaking, individuals, and since individuals as such cannot be caused – although their existence can be caused – causal chains can only *start* with agents. This suggests that agents and their actions can be the beginnings of causal chains, but not intermediate links in causal chains. That would mean that agents can start new causal chains, which will somehow interact with existing ones and influence the future that seemed to be determined. Seemingly, agents and their actions can interrupt the causal chains that existed, and this makes room for the existence of a free will.

If agent causation is to be a reason why free will exists, the step must be made from an agent being the cause of an action to the free will of the agent playing an essential role in this process. The line of argument would then be something like the following:

1. An agent has a free will.
2. This will causes the agent to perform an action.
3. The action was caused by the agent, exercising his free will.

¹¹ This free will libertarianism should be distinguished from libertarianism as a radical liberal political philosophy (Boaz, 1997).

¹² Inspired by the lyrics of the Bob Dylan song ‘The Lonesome Death of Hattie Carroll’.

2.2. Free will and responsibility

Free will matters for responsibility under this view because agents are responsible only if they have free will (O'Connor & Franklin, 2020). John, from our case above, is liable to be punished because he is responsible for not paying attention to traffic and violating traffic rules, and he is responsible because he exercised his free will by turning around, thereby causing the accident that injured Geraldine. John is the agent who caused the accident and he is therefore responsible for the accident. Because he is responsible (which includes being blameworthy), he deserves punishment (and to allow for punishment, criminal liability is imposed). Free will libertarianism, agent causation, and responsibility based on retributivism fit together. We have already seen that human beings have a natural inclination toward retributivism. Similarly, we have a natural tendency to consider ourselves and others as agents with free will (Greene & Cohen, 2004).

This position makes intuitive sense and of the two main justifications for criminal liability, retribution seems the more natural approach. We have an inclination to punish wrong behaviour, because the wrongdoer 'deserved' to be punished. We are psychologically inclined to base criminal liability on the responsibility of the agent for what she did. Given this, why have the positions retributivism entails been increasingly problematized?

3. Arguments Against Retributivist Punishment

The argument against retributivist punishment is often framed in terms of free will and determinism. People can only be blameworthy and responsible for their behaviour if they have a free will. Since – it is assumed in this argument – all events in the universe, including human behaviour, are determined by physical laws, people do not have a free will and cannot be responsible. Criminal liability presupposes responsibility, which presupposes blameworthiness, which presupposes free will, which presupposes the falsity of determinism (Strawson, 1962). If determinism is true, there should be no criminal liability based on retribution.

As indicated in the previous section, determinism holds that any exhaustive set of states of affairs can only be followed by one particular sequence of new states of affairs, meaning that an agent does not have different options available between which she can freely choose. In this section, we first consider determinism in more detail, delineating it from fatalism (3.1 and 3.2), before turning to the arguments against retributivism (sections 3.4 and 3.5).

3.1 Fatalism

Determinism should be distinguished from fatalism, which claims that no matter what action an agent takes, a particular consequence will always result. The difference between determinism and fatalism

is that determinism allows that an action can influence the future. For example, whether Jane buys a ticket in the lottery can make a difference for the question whether she will become a millionaire soon. It is not under her control whether she will buy a ticket, but her action can make a difference for what happens in the future. Fatalism, on the contrary, assumes that your present actions do not make a difference for (some particular aspect of) the future.¹³ For instance, whether Jane buys a lottery ticket or not, she will not soon become a millionaire anyway (see figure 2). Determinism is relevant for the free will issue, as it claims that an agent has at any time only one option open. Fatalism is not relevant for the freedom of the will, because it does not address the issue what an agent can(not) do.¹⁴

Figure 2 assumes that determinism is false: the leftmost circle has more than one possible successor. It represents fatalism as both alternative successors lead to the same final result, which is apparently unavoidable. In the example about the lottery the leftmost circle represents the situation before Jane decides whether to buy a ticket, the two successor circles represent the situations that she bought, or did not buy, a ticket, and the final circle represents that Jane did not become a millionaire.

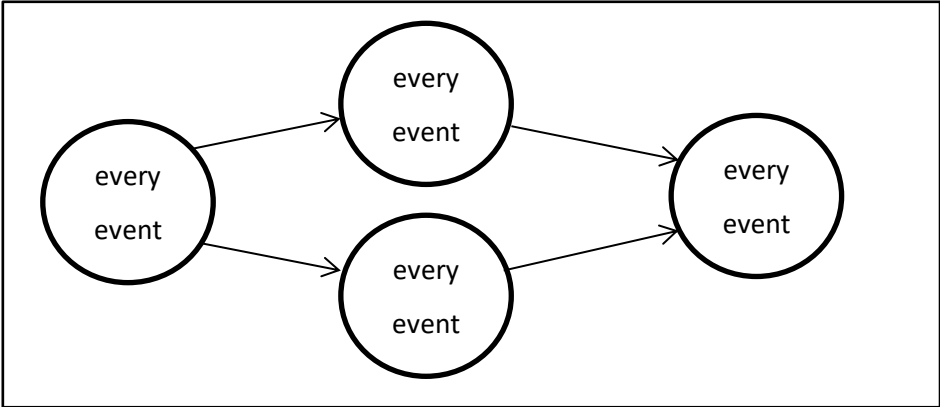


Figure 2: Fatalism

Fatalism as such does not require determinism; it allows that a particular state of the universe has more than one possible continuation. However, all the possible continuations have one or more elements in common and these elements occur whatever an agent does.

¹³ This definition of fatalism differs from the definition given by Bernstein (2002), which reads that every fact, given that it is what it is, could not have been otherwise. In terms of dyadic modal logic: $N(P|P)$. The drawback of this latter definition is that it does not mention actions, while fatalism is a view on the causal inefficacy of actions.

¹⁴ Actually, this is not completely true. Some actions are defined in terms of their consequences. An example would be committing a murder. If the victim will not die, no matter whether the agent fires the pistol, it is not open to the agent whether or not to kill the victim. However, even in cases like this, fatalism does not answer the question whether the agent can fire the pistol.

3.2 What must be the case and what an agent can do

Running ahead of our argument about compatibilism, we want to point out already here that there is a difference between what is necessarily the case concerning an action (e.g. 'it is necessarily the case that Beatrix shoots Henry') and what the agent of that action can(not) do (e.g. 'Beatrix cannot do anything but to shoot Henry'). Determinism holds that given the laws of nature, only one sequence of exhaustive states of affairs can follow a given exhaustive state of affairs. The later states are necessary, given the state from which they originate. Determinism as such does not say anything about what an agent can do. Perhaps this seems a strange remark. Suppose that, at some moment $t+1$, Beatrice decides to pull the trigger and to shoot Henry, and suppose also that determinism is true. Then, given the universe as it actually was just before Beatrice pulled the trigger (at t) and given the laws of nature, it is necessarily the case that Beatrice pulls the trigger at $t+1$? Any fact that occurs after t is necessary, given the exhaustive set of facts at t . Beatrice pulling the trigger is one of them, and that was necessary too: it could not have been otherwise, given the situation at t . So, it seems to follow that Beatrice could not do anything other than pulling the trigger.¹⁵

Yet this does not follow logically. The sentence that at time $t+1$, Beatrice had to (could not do anything but) pull the trigger does not follow from the sentence that it is necessarily the case that at time $t+1$, Beatrice pulled the trigger. The former sentence deals with ability, or the lack thereof, of Beatrice, while the latter sentence deals with Beatrice's actual behaviour and the necessity thereof.¹⁶ As we will point out in Section 4.3, both abilities and necessities are defined in our social practices, and it depends on those practices, not on logic, whether the necessity of the state of affairs that a person acts in a particular way implies that this person lacked the ability to do something else. So, theoretically, it is possible that Beatrice necessarily pulled the trigger (necessity of a state of affairs), but that she could have refrained from doing so (ability to refrain from that action).

¹⁵ The difference between what actions necessarily take place and what an agent can(not) do is central in the debate between compatibilists and incompatibilists that adhere to the so-called consequence argument. See (Van Inwagen, 1995) and the discussion in (Kane, 2005, Chapter 3).

¹⁶ The difference becomes immediately clear if we formalize the example. The sentence that it is necessarily the case that Beatrice pulled the trigger would be formalized as $N(\text{Pulled_trigger}(\text{Beatrice}))$. The sentence that Beatrice had to (could not do anything but) pull the trigger would be formalized as $\sim\text{Could}(\text{Beatrice}, \text{refrain_from}(\text{pull_trigger}))$.

Of course, other formalizations are possible. However, it would be a challenge to find a formalization which *both* does justice to the natural language formulations of the two propositions – the one about what is necessarily the case and the other about what an agent can do – *and* exhibits a logical relation between the two sentences.

3.3 Three positions

The threat that determinism poses to the existence of free will and to the possibility of holding people responsible for what they have done has met with three main kinds of reaction: free will libertarianism, hard determinism, and compatibilism. We have already considered the first, free will libertarianism, above. The second, hard determinism, accepts the determinist threat and its consequences: we cannot hold people responsible for an action if the responsibility is based on the alleged fact that they were free to perform the action or to refrain from it. Either the notion of responsibility should be given up, or – if we want to maintain it – we should use it for consequentialist reasons but give up retributivism. For instance, holding people responsible for what they do, even if they do not deserve it, may lead to better behaviour and a better society (see also Section 5.5). Hard determinism and its implications have recently been advocated by, amongst others, Smilansky (2000, 2002), Peereboom (2001, 2014) and Caruso (2012).

The third kind of reaction is to argue that determinism and free will, or determinism and responsibility, are compatible with each other. This view has become known as ‘compatibilism’, or ‘soft determinism’ (Kane, 2005, p. 22) and has recently been defended by, amongst others, Dennett (1984), Dworkin (2011), and Morse (2013). We will turn to compatibilism in section 4 to examine whether it offers a solution between the conflicting positions of free will libertarianism and hard determinism. In the following, we will first consider the arguments hard incompatibilists bring against free will libertarianism and retributivism.

3.4 Libet’s experiments

Free will libertarianism – at least the strand of it that argues for agent causation – holds that

1. An agent has a free will.
2. This will causes the agent to perform an action.
3. The action was caused by the agent, exercising his free will.

The correctness of this line of argument has been tested empirically by a series of experiments conducted under the responsibility of Libet (1985, 2011). To make this test possible, Libet had to find a measurable correlate of the will, and to that purpose he used the conscious experience of the intention to act.¹⁷ It was already known that an increase of the brain’s electrical activity, called the ‘readiness potential’, preceded voluntary finger-flexing movements (Kornhuber & Deecke, 1965). That neural activity precedes voluntary movement is not very surprising. What Libet wanted to find out is the timing of brain processes in relation to the conscious intention to initiate the movement.

¹⁷ Much of the following text on Libet’s experiments is based on an unpublished text written by our colleague David Roef. We thank David for allowing the use of his text. Of course, we remain responsible for the resulting argument.

The set-up of Libet’s experiment was essentially as follows: subjects were instructed to perform a simple and predefined bodily movement, such as flexing one’s finger or wrist, whenever they felt the urge or wish to do so, while at the same time taking note of the time when the urge or wish to move took place. During this process, EEG-measurements were taken to record the readiness potential. This allowed the experimenters to compare the timing of the onset of the conscious intention with the timing of the readiness potential.

The main finding of the experiment was that the readiness potential precedes the occurrence of the conscious intention to move by about 330ms (milliseconds) and the onset of the movement itself by about 530ms. Interestingly, the awareness of moving also precedes the actual movement in time. The following timeline was adapted from Wegner (2002, 53):

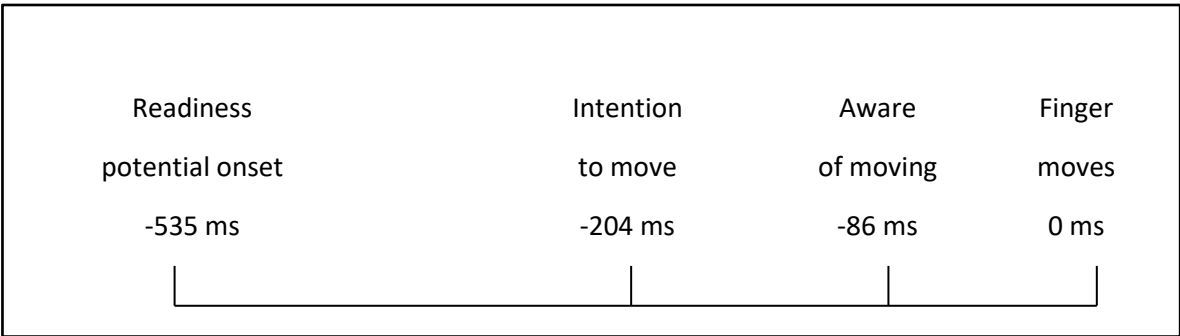


Figure 3: Readiness potential preceding conscious will and action

Libet derived two conclusions from the experiments. The first conclusion is that the conscious intention to move cannot be the *original* cause of action because it comes too late in the neuropsychological sequence. The second claim is that the conscious intention does not initiate the bodily movement, because it could only do this if it preceded, or at least coincided, with the readiness potential (Libet, 1985).

If the activation of the readiness potential plays a causal role in producing an action and if the conscious will to perform this action only sets on after the readiness potential was activated, the conscious will cannot be the original cause of the action or initiate it. This finding seems to overthrow at least a simple view of how a person’s consciousness decides about an action and then as a result of this decision starts a brain process which in the end leads to the muscular events that are an essential aspect of an action.

Libet’s findings are one of the many building blocks of an extended argument presented by Wegner (2002), with the conclusion that human behaviour is not caused by a conscious will. Wegner’s conclusion is that ‘... the experience of conscious willing action occurs as the result of an interpretive system, a course-sensing mechanism that examines the relations between our thoughts and actions

and responds with “I willed this” when the two correspond appropriately.’ (Wegner, 2002, p. 317). If Wegner is correct, the main reason to believe in a free will, namely that we experience having one, is misguided. Perhaps we experience having a free will, but this free will-experience is not an experience of anything outside the experience itself.

3.5 The elusive ‘I’

Libet’s results have received so much attention because they are often interpreted as being relevant for responsibility. This is, for example, reflected in the title of a book on the impact of Libet’s research, which reads ‘Conscious Will and Responsibility’ (Sinnot-Armstrong and Nadel, 2011). People tend to see themselves as agents with a central role for their conscious will that is manifested in the intention to act. They assume that they are persons, typically self-referred to as ‘I’. They also assume that persons are continuous in time, that they have – but are not identical to – bodies, and that they are often – but not continuously – conscious of themselves and of their bodies and surroundings. If a person acts, it is this ‘I’ that acts and if actions are performed intentionally, this ‘I’ has the relevant intention.

Descartes (1641) assumed that this ‘I’ was an independent substance, a ‘thinking thing’ (*res cogitans*). He opposed it to the human body. That would be independent extended thing, a *res extensa*. These assumptions lead to the now unpopular dualism between mind and body and raises the difficult question how the mind and the body can interact. If we reject Cartesian mind/body-dualism, the alternative seems to be to find a place for the mind in the body. There are two main ways to do this, called ‘identity theories’ and ‘epiphenomenalism’ respectively.

Identity theories hold that mental phenomena are identical to states of the body, typically of the brain. Mental phenomena are another aspect of brain states (Smart, 2017) - a mental state is how a brain state is experienced. In some, or perhaps all, of these experiences an ‘I’ plays a role as the subject of the experience. According to identity theorists, it also holds for these experiences that they are the counterparts, in the sense of different aspects, of brain states. If mental states are brain states, they can exert causal influences, such as making a muscle contract, to the same extent that their identical brain states can. Figure 4 depicts this graphically. A circle represents a combined mental/physical state, and the arrows between the circles represent causal relations between these combined mind/brain-states. The lines dividing the circles do not represent a separation between two parts, but a distinction between two aspects of the same ‘thing’.

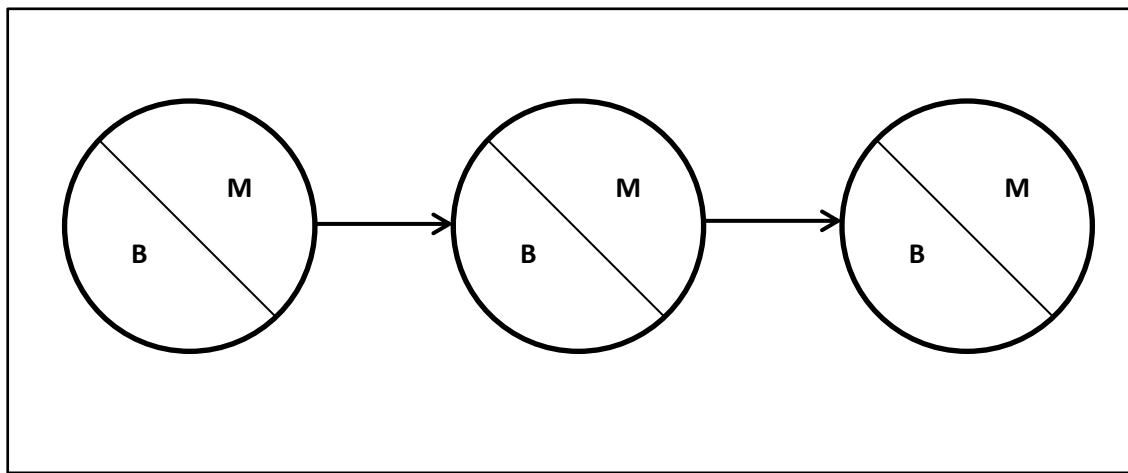


Figure 4: Identity of mind (M) and brain (B)

A main function, perhaps *the* main function of identity theories, is to explain how mental causation (Maslen, Horgan and Daly, 2009; Robb and Heil, 2019) is possible. How can mental states exert a causal influence on the body? The answer of identity theories is that the mental states are at the same time bodily states and that they exert this influence in their role of bodily (brain) states. Assume, for instance, that some brain state is experienced as the desire to drink a beer, that the brain state causes William's muscles to contract, which leads in turn to William's body moving. These effects can also be described as William going to the fridge and taking a beer. Does it make sense to describe this course of events as William's desire to drink a beer causing him to go to the fridge and take a beer? According to identity theorists, this makes sense. According to epiphenomenalists, it does not, because all the causal work is performed by the bodily states, and the mental states might just as well not have existed. Epiphenomenalism is the view that mental events are caused by physical events in the brain but have no effects upon any physical events (McLaughlin, 1994; Robinson, 2019). The relation between brain states and mental states has been compared to that between a steam locomotive and the plumes of steam it ejaculates, respectively. The state of the locomotive at time t influences the state at time $t+1$, and each state goes together with a steam cloud, but the clouds do not influence the movement of the locomotive. Similarly, brain state b_1 at time t_1 goes together with mental state m_1 (also at time t_1). Brain state b_2 at time t_2 goes together with mental state m_2 (at time t_2). b_1 causally influences b_2 , but no causal connections start from a mental state: neither from a mental state to another mental state, nor from a mental state to a brain state. This is depicted in figure 5, where rectangles represent brain states and circles represent mental states. The horizontal arrows represent causal relations, and the vertical lines indicate what mental state goes together with some particular brain state.

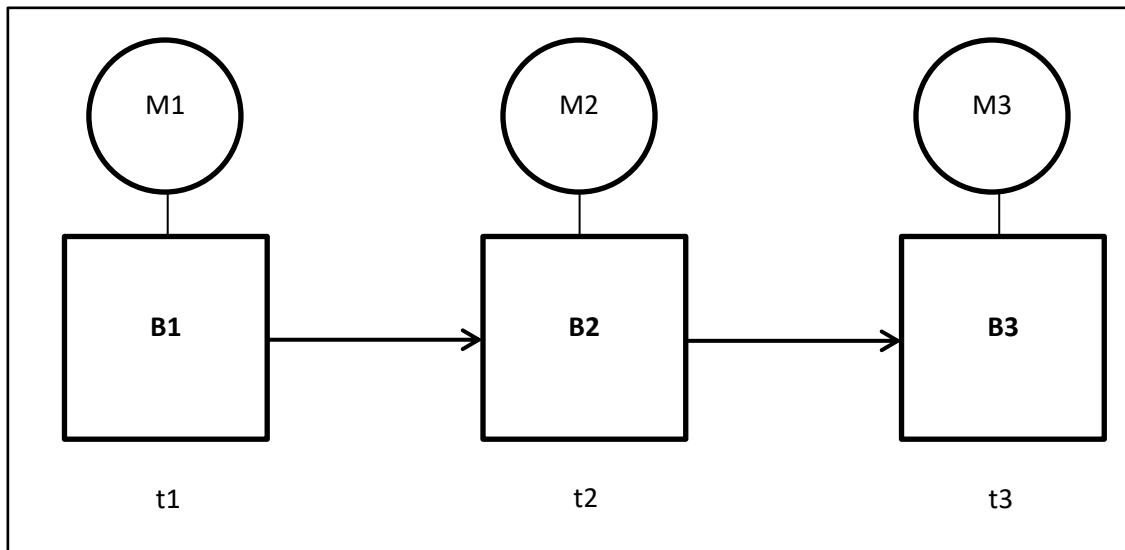


Figure 5: Epiphenomenalism

Epiphenomenalists do not try to find a role for mental causation; they deny that mental causation is possible. If epiphenomenalism is adopted, there is no place for an 'I' as a mental phenomenon in the explanation of actions. Of course, it remains possible to ascribe an action to a person, and to admit that this person experiences herself as 'I', but this is then nothing other than a mere mode of speaking. This 'I' has no place in the causal processes that lead to actions or other events. For the same reason, intentions as mental states have no causal influence. The events that are called actions may be accompanied or preceded by mental events such as intentions to act, but these intentions play no causal role in the generation of the actions.¹⁸ Libet's experiments are fully compatible with this epiphenomenalist view.

If it is assumed that mental states are identical to bodily states – for instance brain states – the prospect for finding an 'I' that acts and that initiates bodily movements seems better. It is 'only' necessary to find a part of the brain, or a brain state, that embodies this 'I' and that realises conscious experiences, including intentions to act, and that initiates intentional actions. The problem is that such a brain part or brain state has never been found. Moreover, the Libet experiments seem to show that there are at least some intentional actions that are not initiated by a brain state that also embodies the relevant intention to act. If there are *some* intentional actions that are not initiated by a brain state that also embodies the intention, it seems at least possible that *no* intentional action is initiated in that way.

¹⁸ Contrary to what is claimed in (Searle, 1983, pp. 83f).

3.6 From free will libertarianism vs. hard determinism to compatibilism

The libertarian theory of agent causation requires that there is an agent, an 'I', from which actions originate. This very notion of an 'I' that can represent a person or other agent is problematic. If an 'I' is identified with a substance that is conscious, there is a problem with the discontinuity of consciousness. Does the 'I' stop existing if a person is dreamlessly sleeping, or in a coma? If it does, how does the 'I' of an earlier stream of consciousness relate to the 'I' of a later conscious experience? Is it the same 'I', and if so, what makes it the same (Strawson, 2011; Metzinger, 2011)? Questions like these are hard to answer, and perhaps they do not even make sense and are 'empty questions' (Parfit, 1984).

Nevertheless, most people tend to experience themselves as continuous in time, and as agents who decide about their actions, who plan and perform many of those actions intentionally, who have reasons for what they are doing and who let these reasons influence what they do. These experiences underlie our social practice of holding people responsible for their actions and making them liable for (some of) the consequences. The problem seems to be that these experiences and the social practices that build on them are hard to reconcile with the results of science and the philosophical theories that elaborate on these scientific insights. In Section 5, we will return to this apparent conflict, but before doing so, we want to turn to compatibilism, the third position in the debate surrounding responsibility, determinism, and free will.

4. Compatibilism

The philosophical discussion about determinism and free will is highly technical. The notion of determinism needs to be defined in terms of states of affairs or events (or a combination of both) and invokes the idea of necessity, which itself is not very transparent either. The phrase 'free will' has only a limited use in common parlance. Moreover, this use – e.g. 'She did this out of her own free will' – stands for voluntariness, which is not the same as the philosophical notion of free will (Van Inwagen, 1995). This raises the sceptical question whether the philosophical discussion has much practical relevance, and even whether it should have such relevance.

This approach, which starts from the question of why we are bothered by determinism, takes a central place in the work of Dennett (1984, 2003). Dennett lists a number of reasons why people may be worried about determinism. They include the ideas that if determinism is true, we are like prisoners robbed from the possibility to do what they want, or like persons whose will is manipulated by a nefarious neurosurgeon, or that people are otherwise the playthings of external forces, or like automata who merely mechanically follow behavioural patterns imprinted in us, and so on (Dennett, 1984, pp. 7-12). Slightly less evocatively, but more systematically, Kane lists objections that

compatibilists adduce against determinist critics of free will. He does so under two headings: the determinist critics misrepresent 1) freedom as absence of determinism, while it should be absence of (relevant) constraints; 2) determinism as a constraint, while it is a form of necessity.¹⁹

Compatibilism about determinism and free will has its basis in this scepticism about determinism as a criticism of free will and responsibility. In one variant, compatibilism holds that it is reasonably possible to hold somebody responsible for her actions, even if she lacks a free will. In a second variant, it holds that an agent can have a free will, even if determinism is true (McKenna and Coates, 2020).

To substantiate compatibilism, it is possible to refer to our social practices, as they are known to us through our intuitive judgements about being free or responsible. These practices may be moral, and then the determinist criticism of free will and responsibility should be tested against our moral intuitions. The practices may also be legal, in which case our law is the touch stone for evaluating theories about free will and responsibility. In the two following subsections, we will briefly discuss two examples. The first one deals with the relation between determinism and moral responsibility; the second one with the relation between, on the one hand, determinism and free will and, on the other hand, legal responsibility and liability.

4.1 Frankfurt on moral responsibility

In a series of articles (a.o. Frankfurt 1969, 1971 and 1982), Frankfurt developed a theory of freedom of the will. In this theory, he tried to overcome the problem that determinism poses for the requirement of alternate possibilities. We have already seen that the availability of alternate possibilities for action is one of the two requirements that are often assumed for the existence of a free will, the other requirement being that the action an agent performs is 'up to her'. Frankfurt claims that the requirement of alternate possibilities can be dropped and replaced by the requirement that a person not only wills her action, but also has a second-order volition that she wills what she first-order wills. If this requirement is met, the person wholeheartedly endorses her action, and then she can be said to have acted out of her free will.

For instance, if an alcoholic drinks another bottle of gin and wanted to do so, his addiction may have stood in the way of his free will. The alcoholic did not want to want another bottle of gin, although he

¹⁹ *Digression on necessity and constraints.* Arguably, all necessity is the result of *some* constraint(s). For example, the sentence 'It does not rain if it is not the case that it rains' is made necessarily true by the constraints of logic. Or, it must be the case that a particular bachelor is unmarried because of the constraints of semantics. (Hage, 2016) However, it does not follow from the fact that something is necessarily the case, without a gloss referring to a point of view, that it was constrained *in a particular way*. Moreover, a particular constraint may make that something is necessarily the case, or that something necessarily happens, but the constraint is not identical to this necessity. The necessity of a particular logical truth is not the same as the logical rule that makes it necessary.

actually (first-order) wanted to drink another one. If this second-order volition is lacking, the alcoholic does not wholeheartedly want to drink another body of gin. In a sense he is forced to do so by his addiction. The force is in this case a force that influences his first-order will, which is therefore not free. However, if an alcoholic completely endorses her alcohol-centred life and wants to want to drink more, her will is free. In that case, it is wholly up to her if she drinks another bottle of gin. So, where originally it was assumed that there are two requirements for the existence of a free will, namely the existence of alternate possibilities and that her actions are 'up to the agent', now there is only one requirement left. This requirement, that a person wholeheartedly accepts her own will, can be seen as an alternative formulation of the requirement that an action is 'up to the agent'.

Let us ignore the question whether Frankfurt's analysis of free will is correct,²⁰ choosing to focus on the method that Frankfurt implicitly applies instead. This method is to test imaginary cases against the reader's intuition. The example about the alcohol-addict is a case in point. How do we know that the ordinary addict lacks a free will, precisely because of his addiction? We have no sense to detect a free will if we encounter one. Somehow, we just 'know' that the will of addicts is not free and that their addictions compel them to 'will' things that they do not 'really' will. Moreover, we know in the same way that the typical addict does not 'really' (second-order) will to satisfy his addiction.

Not only do we know this, but we also know that others know the same. We would be surprised if others claimed that an ordinary addict, one who has not voluntarily chosen to be addicted, is happy with his first-order desires. It is part of the very notion of an addict that he does not endorse his first-order desires; the possibility to refer to such people who lack the relevant second-order desires is the very purpose for which we have the term 'addict' in our vocabularies (Hage, 2018, pp. 39/40). We do not need empirical enquiries and we can normally rely on our intuitions to determine whether a person has a free will. This is so, we venture, because during our socialisation process we have been introduced to a social practice in which we distinguish between persons with ordinary volitions and persons who are addicted, and in which we distinguish between persons with and without a free will. So, when we use our intuitions to test theories about when free will is present or absent, we try to systematise a social practice which is already present.

In this chapter, we have used Frankfurt's approach to the free will discussion as an example, both because Frankfurt's view has been very influential on the free will discussion, and also because his work is illustrative for much research that is done on moral responsibility and its requirements. The method that Frankfurt employs is used in much research on free will and responsibility, in particular in

²⁰ There are many papers addressing this issue. For an overview of some of them, see Kane 2002, pp. 281-334. One issue is whether second-order volition suffices for the will to be free. Is it not also necessary to have similar third- and higher-order volitions? This discussion is similar to the discussion of infinite regress in connection with constructivist facts. See, Hage (Present Volume).

research that leads to compatibilist outcomes. Frankfurt formulated theories about the nature of free will and tested these theories against our intuitions about specific cases. These intuitions are more or less taken at face value. If the research method is based on systematizing moral intuitions, it is not done to question these intuitive judgements on a large scale. Researchers explore our social practice as it actually is, without proposing far-reaching alternatives for how we should see responsibility and free will in our daily lives. The following presentation of the views of Morse provides a good illustration of this with regard to law.

4.2 Morse on legal responsibility and liability

In law, the relevant social practice to analyse is positive law itself. The question of how to see determinism, free will, responsibility and liability in law is answered by the law. In a relatively recent article, Morse has done the relevant study for the criminal law in common law systems (Morse, 2013). Here we will mostly confine ourselves to quoting phrases from the article in question and adding some comments. All the italics in the quotations are the present authors' (JH & AMW) and footnotes are omitted.

'The state may imprison people in the criminal justice system if they *deserve punishment* for crimes they have committed, and it may civilly commit dangerous people if they are not responsible agents [...]'

'Virtually all criminals are rational responsible agents, and according to the dominant story, the deprivation imposed on them – punishment – is premised on considerations of desert.'

Here we see a sharp distinction between criminal and private law. Criminal law is deemed essentially retributivist, while private law may be based on consequentialist considerations, such as prevention. Moreover, the reason why criminals are punished is that they deserve punishment, and they deserve punishment because they are rational and responsible agents. Here we encounter the idea of reason-responsiveness that is by many considered the ground for holding people responsible for what they do and, in criminal law, for punishing them.

'The law presupposes the folk-psychological view of the person and behaviour. [...] *folk psychology considers mental states fundamental to a full explanation of human action*. Human behaviour cannot be adequately understood if mental state causation is completely eliminated.'

'We are practical reasoners, the sort of creatures that can act for and respond to reasons. *The law does not treat persons generally as nonintentional creatures of mechanical forces of nature*. As a

system of rules and standards that guides and governs human interaction, law tells citizens what they may and may not do [...]

The concepts that are used in criminal law are those used by folk-psychology. They deal with mental states, and not with, for instance, neurons or readiness potentials. Morse claims that these folk-psychological concepts are essential for a proper understanding of human behaviour. This claim is supported by the function of law which addresses people as users of guiding reasons (rather than as objects determined by physical forces). This is, again, the idea of reasons-responsiveness.

'[...] if the agent lacks a requisite mental state, the agent is not also prima facie criminally responsible and must be acquitted outright of the crime requiring that mental state.'

'Mens rea terms, such as intention, purpose, knowledge, recklessness [...], and negligence [...] have their ordinary language, common sense meanings. No degree of commitment or rationality is included in the definitions of mens rea.'

Mens rea is an essential condition for most, if not all, crimes, and the different forms of *mens rea* are psychological situations.²¹ There is no (implicit) reference to the possession of a free will. Moreover – and we do not include quotations to illustrate this, but Morse claims it – the affirmative defences through which a defendant can prevent punishment even though he is prima facie responsible for a crime, are defined by the law in plain categories such as insanity, or self-defence, without reference to philosophical doctrines about determinism, or actions 'being up to the agent'. This suggests that the free will discussion as it is conducted in the philosophical literature is not relevant for the law as it actually is.

Is this something we should be willing to accept? According to Morse it is, because

'[...] ordinary people are intuitive compatibilists, and my hunch is that even the most committed hard incompatibilists actually live as if they were compatibilists. Finally, [...] it seems clear that interpersonal life would be exceptionally impoverished if concepts of responsibility, including genuinely deserved praise or blame, were extirpated from our lives, even assuming that this is

²¹ It is open to discussion whether negligence is really a mental condition. Perhaps it is better described as a kind of unlawful behaviour that was performed unintentionally. With some good will, this lack of intention may be seen as the mental condition that satisfies the *mens rea* requirement. Cf. (Blomsma and Roef, 2019a, p. 195).

possible. If people came to be treated as objects to be manipulated rather than full agents, much that we most treasure would be lost.'

Apparently, our actual social practice is so attractive that we should stick to it, irrespective of whether it is based on a sound image of humankind. It is what it is, and it is good (enough).

Morse's argument contains two main lines. The first line relates the legal reasons on whether to hold people responsible for their behaviour to the reasons in the philosophical discourse and finds that they are by and large disconnected. The law does not presuppose the absence of determinism, the presence of a free will, or an action 'being up to the agent'. To the extent that the law implicitly works with these notions, it does so by assuming that these conditions are fulfilled, unless the contrary is shown by means of one of a limited number of exceptions that are explicitly allowed by the law. Legal debates do not deal with the philosophically big issues, but with the exceptions allowed by the law such as absence of *mens rea*, or presence of insanity, or duress.

The second line deals with the (im)possibility and undesirability of abandoning our actual social practice. Morse admits that our actual social practice of punishing those who committed crimes for the reason that they deserved punishment is not yet perfect. However, he does not see any fundamental flaws, but merely imperfections in what is basically good. So, there is no need for large-scale modification of our practice of criminal law, and Morse suggests that, even if this were needed, it would be impossible to accomplish. This suggestion echoes the views of Peter Strawson (1962), who claimed that we are not able to replace our actual social practice of holding people responsible and having moral emotions by an instrumental one which treats people as elements of physical reality rather than as responsible persons.

4.3 The practice of agency

Compatibilists typically take as their starting point our present social practice of holding people responsible for their actions. This practice makes no claims about big philosophical issues such as determinism, free will, and actions 'being up to the agent'. Rather, it deals with common-sense notions such as intellectual disability, insanity, addiction, duress, momentary confusion, etc. Although it is possible to relate these common-sense notions to the philosophical issues, we tend not to do so in our daily practices. We typically do not excuse people for the reasons that they lacked a free will, or that their behaviour was determined by past events and the laws of nature. We do so because they acted under duress, or that they are mentally handicapped. This difference creates the impression that our actual practice is compatible with any answers to the big philosophical questions. Is this impression correct? We will discuss this question by means of two examples relating to determinism.

The case of Esther. Esther received a good education from her parents. One of the things they told her is that it is wrong to discriminate against people because of their ethnic or religious background. They informed Esther that history contains several examples of peoples that were slaughtered precisely because of such discrimination, and that there were dictators who encouraged or even initiated this. It would have been better if these dictators had never lived.

Esther took her parent's words to heart. When she turned thirty, a racist politician rose to power in her country. Esther decided that it would be better to kill the potential dictator now, while it was still possible, rather than wait until the politician had enough power to become invulnerable and start a genocide against people with Esther's ethnic background. Esther shot the politician dead, was arrested, and brought to trial.

Esther's advocate argued that Esther should be excused of the crime of murder because of her upbringing and because of her fears. The court did not accept this defence, however. It argued that, apart from the killing, Esther was a normal person, and that an anti-racist upbringing or an unjustified fear for genocide could not excuse a political murder. Neither Esther's advocate, nor the court mentioned anything about Esther being determined (in the philosophical sense) to kill the politician.

The case of Harry. Harry was an ordinary boy until his early twenties. At that age he started to become afraid and suspicious of what people in his environment were doing. He consulted his physician, who referred Harry for psychiatric treatment. However, the mental health services in Harry's country suffered from underfinancing. They had a large backlog of work, and the waiting time for even having an intake conversation with a qualified mental health care worker was more than a year.

During this year, Harry's condition worsened considerably. He came to believe that a politician who rose to power in his country would become a dictator and would start a war against neighbouring countries. Harry decided that it would be better to kill the potential dictator now, while it was still possible, rather than to wait until the politician had enough power to become invulnerable and start these wars. He shot the politician dead, was arrested, and brought to trial.

Harry's advocate argued that Harry should be excused of the crime of murder because of his mental condition. The court accepted the insanity defence, ordered that Harry would receive treatment for his mental illness, but did not penalise Harry for the murder. Neither Harry's advocate, nor the court mentioned anything about Harry being determined (in the philosophical sense) to kill the politician.

These two cases have in common that a set of circumstances brought a person to commit a political murder. In Esther's case, these circumstances were not recognised as a valid excuse and Esther was convicted for the murder. In Harry's case, the circumstances were recognised as an excuse and Harry was not convicted, although he was committed for medical treatment.

What circumstances count as an excuse is determined by the legal system in question, which in this respect often follows pre-existing social practices. People who commit crimes for merely putative justifications are typically not excused, while people who do so because of what is recognised as a mental illness are excused. That is how society operates, and the law reflects this social practice.

As compatibilists argue, these social practices do not refer to the philosophical issue of determinism but deal on a case-to-case base with the circumstances that may constitute an excuse. Non-compatibilists may argue that implicitly, these social practices nevertheless reject determinism. What counts as an excuse may be based on an implicit view on what an agent can or cannot avoid. Apparently, society assumes that persons like Esther can refrain from political murders, while persons like Harry cannot.

The hard determinists claim, however, that determinism – or, rather, the facts and physical laws to which the doctrine of determinism refers – makes that nobody can help anything they do. Society wrongly distinguishes between Esther's and Harry's case: if there is a reason to excuse Harry, there is also a reason to excuse Esther.

The compatibilist can rebut that society does not consider determinism relevant for what an agent can do. Indeed, underlying the justification of criminal sanctions is a theory of what agents can do or avoid. This theory, so claims the compatibilist, does not depend on whether determinism is true; it is also part of our social practice (Hage, 2016).

The picture that arises is that of a complex social practice, which uses a set of interrelated concepts, such as action, agent, intention, negligence, blameworthiness, justification, excuse, ability, necessity, responsibility, liability, cause, effect, voluntariness, force, and duress. However, this practice does not deal with free will or determinism. We will call this social practice the *practice of agency*, because of the central role that actions, agents, and responsibility play in it. The practice of agency defines its own concepts and, as compatibilists argue, it does so with little regard for big philosophical issues such as determinism, free will and the foundation of responsibility. Because the concepts used in the social practice of agency are defined in this practice, independent of the philosophical issues, the philosophical debate and the social practice seem worlds apart. The social practice uses notions that, in the eyes of philosophers, only make sense against the background of determinism and free will. Insanity or duress are in relevant the eyes of these philosophers because they are signs of determination and lack of free will. However, these same philosophers neglect the issues that our social practice deals with. They are lost in the 'big issues' such as determinism and free will and overlook that in society, the 'minor' topics such as insanity or duress are more relevant than these bigger issues. This conceptual distance gives rise to the belief that our practice of agency is compatible with the philosophical theories, even if these philosophical theories are inconsistent between

themselves. This would mean that the practice of agency can be continued without worrying about the issues raised by some philosophers. Compatibilism, it seems, is able to reconcile retributivism and responsibility based on retributivism, and the philosophical arguments.

That conclusion does not follow, however. Even if adherents of the existing social practice, such as Morse, could show that their practice is not vitiated by philosophical theories, this leaves open the issue whether the practice of agency as such is justified. Perhaps, Peter Strawson is right in assuming that we are not able to abandon our practice of agency and perhaps Morse is right in claiming that our practice is not so bad that it requires to be abandoned. However, those are as yet unsubstantiated assumptions.

5. Is Compatibilism the Solution?

The central question which this chapter has wrestled with is whether criminal liability can and should be based on retribution. The previous sections have discussed this question from the perspectives of three main positions in the debate. The first position, free will libertarianism, rejects determinism and holds that free will exists, that agents are responsible and deserve punishment for wrong actions, and that criminal liability based on responsibility is possible. The second position, hard indeterminism, is that determinism makes free will impossible, that free will is a necessary condition for responsibility and criminal liability, and that therefore the truth of determinism makes criminal liability based on responsibility impossible. The third position is compatibilism. Its many variants share the idea that the philosophical debates on determinism, free will and responsibility deal with other issues than our actual social practice, that the outcome of the philosophical debate therefore does not affect this practice and that determinism and the absence of free will are therefore compatible with criminal responsibility and liability.

Our aims in this chapter were to present the main lines of the debate surrounding responsibility and free will, given their relevancy for the question whether criminal liability should be based on responsibility and responsibility on retributivism, and to examine whether the different positions within the debate can be reconciled. So far, we have considered free will libertarianism and hard indeterminism as two opposing sides of the debate, and compatibilism as a possible form of reconciliation between the two. In the following section, we want to examine whether compatibilism can succeed at this reconciliation.

5.1 Internal and external questions

In order to examine whether compatibilism succeeds in reconciling the different positions of the debate, it is necessary to gain more distance from the details of the main positions, in order to obtain

a suitable perspective on the continuing debate. In a conceptually illuminating article, Mackor (2013) distinguishes between four kinds of questions that can be asked about the relevance of the cognitive sciences for legal responsibility. There are two 'internal questions'. One concerns the issue whether the neurosciences can change the conditions that our social practice uses for assuming responsibility. For instance, should we still consider addiction as a reason why the addict is not responsible? The other addresses the issue whether the neurosciences can be of help in answering the question whether in a particular case the conditions, as they are actually used, have been fulfilled. For instance, did the brain tumour actually make it impossible for the defendant to act on his knowledge of right and wrong?

There are also two 'external questions', which deal with the viability of our social practice of holding people responsible as a whole. One of them concerns the issue whether our social practice is based on false presuppositions, for instance the presupposition that agents have a free will. The other addresses the question whether our social practice promotes its own purpose, for instance the purpose of preventing crime.

The questions addressed in this chapter are typical external ones. The existence of free will and responsibility are allegedly presuppositions of our practice of agency, although compatibilist deny this. The question whether our practice of agency is worthy to be rescued from the criticism may be answered on the basis of the purpose of this practice, and is also external. This section returns to these questions as questions about the tenability (presuppositions) and the desirability (purpose) of our actual practice of agency. We will start with another look at the presuppositions by distinguishing the phenomenological and the realist view of the world.

5.2 The phenomenological and the realist view

When we grow up, we learn to live in a meaningful world. Already at a young age, we identify objects in the world around us, and these objects have colours and smells, they are beautiful or ugly, quiet or noisy, hard or soft, kind or unkind, heavy or light, etc. In this first stage, we do not distinguish between what and how the objects around us are and how we experience them: they are assumed to be as we experience them.

At some moment, we start to distinguish between the objects and their characteristics as they are 'in themselves' and how they appear to us. The stick that seemed to be bent when standing in water turns out to be straight if taken out of the water. The colours that we see turn out to be different for colour-blind people. Moreover, it becomes possible to raise philosophical questions, such as whether a tree that falls in a forest without living beings around makes a sound.²²

²² See the discussion in the Wikipedia encyclopaedia: https://en.wikipedia.org/wiki/If_a_tree_falls_in_a_forest

The insight that we can distinguish between how things are 'in themselves' and how they appear to us is not new. A famous example from antiquity is the cave-simile that Plato presented in his dialogue about Justice and the State (The Republic, 514A-520A). The simile tells us about people who are fettered in a cave, where they can only see shadows of real things. One of them manages to escape from the cave and learns what the real world, illuminated by the sun, looks like. When he returns to the cave and tells the others what the real world is like, the others do not believe him and ultimately, they even kill the enlightened adventurer. Today, scientists who tell us that the real world is very different from how it appears to us do not run a big risk of being killed, but the resistance against their views takes different forms, including the opinion that for our daily lives, we can ignore their findings. Let us call the view that assigns a central role to how we experience the world, the 'phenomenological view' (Smith, 2018). This phenomenological view can be opposed to the view that aims at knowledge of the world as it is in itself, independent of our experience or knowledge of it. Let us call this the 'realist view', borrowing the term 'realist' from the realist current in philosophical ontology (Miller, 2019). Unavoidably, human knowledge starts from how we experience the world. This is the perspective with which individual human beings grow up. Arguably, a similar development has taken place in the development of humankind, as is evidenced in Plato's cave simile. Much of modern science aims at collecting and systematising knowledge about the world as it exists independent of our experience of it. Our mental phenomena, including our mental states and including the meanings we attach to the world that we experience, change from being a window to the world into objects of knowledge, analogous to the objects in the 'real world' (in the philosophical-technical sense). Rather than taking the world-as-experienced, our 'life-world' in phenomenological jargon (Gallagher, 2012, pp. 159-164), for granted, as happens in the phenomenological view, the realist view of the world tries to account for the world as it really (mind-independently) is.²³ As the experiences that are caused in human perceivers by the external world are also seen as parts of the real world, they become phenomena that need explanation, just like other parts of the real world. Realist science aims to explain the external world as well as the experiences created by this world, and the relations between them (e.g. how can we explain vision?).

The realist approach can also be taken to our experiences of ourselves. For Descartes, a person's experience of himself as thinking was sufficient evidence to derive that there was a thinking substance. Logically speaking, however, it is impossible to derive the existence of a thinking substance from the existence of an experience with as propositional content 'I think'.²⁴ As there is no clear other support

²³ We ignore, for the purpose of this paper, questions of philosophy of science relating to whether it is possible to account for the world as it really is, how much theory is involved in scientific enterprises, etc.

²⁴ For logicians: It is not possible to derive from $\exists e(e = \text{experience}(\langle \text{jaap thinks} \rangle))$ that $\exists p(p = \text{jaap} \ \& \ \text{Thinks}(p))$. The text between '<' and '>' represents the propositional content of the experience. Cf. Quine 1956.

for the existence of an 'I' apart from the I-aspect of our conscious experiences, the realist approach to human existence may be the cause of the vanishing 'I' that was discussed in section 3.5.

In section 3, we encountered the argument that if determinism is true, there is no free will because of a lack of alternate courses of action. If determinism is false, there is no free will either, because then it is not 'up to the agent' what course of action she will take. So, apparently, the absence of free will follows from both determinism and indeterminism and the determinism issue is not relevant for the existence of free will. Nevertheless, it is the discussion of determinism that shows us why there is no free will. At this stage in our argument, we can explain why determinism is irrelevant to the issue of free will – no matter whether determinism is true, there is no free will – while at the same time showing why the discussion of determinism is relevant. The discussion of determinism and its relevance for free will takes the free will issue out of the phenomenological domain and puts it in the realist domain. In the realist discourse, the phenomenological ability to choose what course of action to take becomes the realist possibility that the development of the world in time takes a particular direction. As soon as this shift in the kind of discourse has taken place, free will has become problematic, either because of a lack of alternate actions, or because the agent does not determine the course of events (or because there is no agent at all). The recent results of the cognitive sciences are therefore not relevant for the existence of free will in the sense that they show that free will does not exist. They are relevant because they are an extra step in the discourse shift from phenomenological to realist.

5.3 Reductionism

The effects of this step from phenomenology to realism are further strengthened by a development in the sciences. The sciences strive for explanatory parsimony by trying to derive more specific laws from more general ones (Curd and Cover, 1998, pp. 903-1047). For example, theories about the motion of the planets are derived from a general theory of gravitation, which is in turn derived from the general theory of relativity and the curvature of space/time. This development goes hand in hand with a reduction of the entities in more specific and concrete theories to entities in more general and more abstract theories. The entities in modern physics become increasingly further removed from the entities in the meaningful world in which we live.²⁵

This development towards reduction of the elements of our 'life-world' to elements that are so far removed from our experienced world that they have become unrecognisable makes it difficult, if not impossible, for people to live in the world as physics describes it. Herein lies the cause of the alleged impossibility, pointed out by Peter Strawson, that we live in a world in which human beings are treated

²⁵ Moore (2018) discusses different varieties of reductionism and their relevance for responsibility.

as mere objects. (Apparently, he still assumed that the world about which physics talks contains human beings, rather than collections of subatomic particles, or only 'fields'.) The replacement of the discourse about the human 'life-world' with a discourse about the entities in the latest physical theories may well be impossible for most human beings. Moreover, as Morse pointed out, such a replacement may go at the cost of what we highly value in inter-human relations.

5.4 The possibility of compatibilism

This supports the compatibilist solution: why not use the two kinds of discourse side by side? Apparently, we can both talk about the chemical reactions that make our muscles contract, and about one person catching another person who almost fell. These are just two perspectives, used to describe the 'same' event in the world. If physics prefers the one level of description, while humans prefer for their mutual relations the other level, nothing seems to be wrong.

A popular metaphor in this connection is that of different kinds of maps that represent the same area, for instance a map of waterways and a map of industrial activities. The different maps give us different information about the same area, and the information of the one map cannot be reduced to the information provided by the other map, but one could be superimposed on the other or they could be combined into a new, more detailed map without contradiction. As long as the information provided by the maps is consistent, there is nothing wrong. However, if the maps contain inconsistent information, for instance because the one map paints a lake exactly where the other maps paints an industrial plant, roads included, they cannot be used together, and there is good reason to assume that at least one of the maps must contain errors.

The same holds for different perspectives on action. If the one perspective provides us a story about chemical reactions in a pack of neurons, while the other perspective talks about a person pondering a difficult decision, it must be possible to combine these two perspectives if they are to deal with 'the same thing'.

However, the consistency of two perspectives on the same thing must allow the perspectives to be perspectives on 'the same thing'. This kind of consistency, which deals with different aspects of the same thing, is not the same as the consistency of theories that deal with different topics and which for that reason cannot contradict each other, like two maps of entirely different areas cannot contradict each other. Another example of the latter would be the way in which theories about the opening times of cinema's in Paris and about the best way to maximise the life-time milk production of a cow are consistent. These two topics have nothing to do with each other, and that is precisely the point of this example: theories that deal with topics that have nothing to do with each other will typically be consistent in that they do not (and cannot) contradict each other.

It is more difficult to achieve consistency of theories that deal with ‘the same thing’. Yet, that is the kind of consistency that is required for the tenability of compatibilism. A neurophysiological theory should be consistent with a theory about belief formation, or about the choice for an action. More is required, however: if compatibilism is to become plausible, the neurophysiological theory and the folk-psychological theories should not merely be consistent with each other about belief and action, it should also be made plausible that they are theories of ‘the same thing’. This requires either that the entities in the phenomenological/folk-psychological account can be reduced to entities in the realist account, or that there is a neutral third way of identifying entities into which both accounts can be translated.²⁶ Moreover, ideally it should also be possible to explain the folk-psychological theories from the neurophysiological ones²⁷: why do particular brain states and –events lead to particular beliefs or actions?

We find that the contrast between, on the one hand, a phenomenological/folk-psychological perspective on human action – which we see reflected in free will libertarianism –, and, on the other hand, a realist/reductionist perspective – which we see reflected in hard determinism – cannot easily be overcome by stating that the perspectives talk about different things and are therefore consistent and compatible. A successful compatibilist reconciliation of the two perspectives asks for more, including at least an account of how the two perspectives are perspectives on ‘the same thing’. Such an account is at present still lacking, and part of the worries of non-compatibilists is that such an account will turn out to be impossible. Whether this worry is justified is still an open question. Impossibility is hard to prove.

5.5 The desirability of compatibilism

Either compatibilism is tenable in the sense that it can be made plausible that the phenomenological/folk-psychological account of mental phenomena and actions deals with ‘the same thing’ as the realist/reductionist account, or it is not. In both cases, it is possible to plead for the co-existence of the two kinds of discourse, be it that the plea is (even) more attractive if compatibilism is tenable. However, if we assume that there is no place for blame or responsibility in the realist/reductionist account, the continuation of the phenomenological/folk-psychological account cannot be based on the fact that it best fits with the blameworthiness and responsibility of agents. If the traditional account is to co-exist with the modern scientific account, it must be justified on non-

²⁶ For lawyers: this is the philosophical counterpart of the demand for a *tertium comparationis* in the theory of comparative law (Michaels, 2006).

²⁷ Theoretically, it would also be acceptable if the neurophysiological states and events could be explained from the folk-psychological ones, but that seems (even more) unlikely.

retributivist grounds. In practice, this would mean that we need a consequentialist justification for retribution in criminal law.²⁸

Rawls (1955) famously argued that the social practice – he called it an ‘institution’ – of criminal law should be justified through its positive consequences. These might, for example, be the prevention of criminal behaviour through incapacitation or deterrence, reparation of the harm done through crimes, or the rehabilitation of criminals. At the same time, individual cases of punishment could be justified by retributivist reasons: the rules defining criminal liability may require blameworthiness as a condition for criminal liability. This would fit with the practice described by Carlsmith (2008) that people who must justify punishment in general often do so by means of consequentialist reasons, while actual punishing behaviour seems to be motivated by retributivism.

The Rawlsian solution seems to provide an attractive way out of the retributivist/consequentialist dilemma in criminal law. It remains to be seen, however, whether this way out really works. That it is theoretically possible to give a consequentialist justification for a retributivist social practice does not yet prove that this consequentialist justification can actually be given. Whether that is the case depends on the actual consequences of retributivism. And precisely here, we can count on the cognitive sciences for new insights.

6. Conclusion

In this chapter we have considered the debates in criminal law theory about responsibility and liability, and how the use of these notions is threatened by determinism, or by the lack of agent control. These debates have been mainly philosophical, although they have been influenced by the upcoming physical sciences with their increasing explanatory power and, quite recently, findings from the cognitive sciences, including the experiments of Libet.

The main positions in these debates have been (1) libertarianism, which defends a central role for agents and their free will, (2) hard incompatibilism, which denies the existence of a free will and the possibility of responsibility or liability based on desert, and (3) compatibilism, which argues that our actual social practice does not depend on any philosophical position and can therefore co-exist with both libertarianism and hard incompatibilism.

We have tried to describe the different positions in the debate in a fair manner. Moreover, in the penultimate section we have tried to explain the continuing debates by means of a distinction between the phenomenological/folk-psychological way of looking at the world in general and agency in particular, and the realist/reductionist perspective. In doing so, we have reformulated the core

²⁸ It is somewhat akin to an atheist pleading for the promotion of (civil) religion because it has such attractive consequences (cf. Rousseau, 1762, book IV, Chapter 8).

business of compatibilism. The central question is not anymore whether free will is compatible with determinism, or whether responsibility is compatible with the non-existence of free will. It has become whether, and to what extent, the realist/reductionist perspective on the world and on agency is compatible with the phenomenological/folk-psychological perspective. The cognitive sciences operate precisely in the border area of (folk-)psychological phenomena that need to be explained, and realist theories that are often used to provide the explanations. It seems, therefore, that the successfulness of the cognitive sciences will determine the successfulness of this new form of compatibilism.

References

- Bernstein, Mark. (2002). Fatalism. In *Kane 2002*, pp. 65-81.
- Blomsma, Jeroen & Roef, David. (2019a). Forms and Aspects of Mens Rea. In Johannes Keiler & David Roef (Eds.), *Comparatible Concepts of Criminal Law*. Cambridge: Intersentia, pp. 177-205.
- Blomsma, Jeroen & Roef, David. (2019b). Justifications and Excuses. In Johannes Keiler & David Roef (eds.), *Comparatible Concepts of Criminal Law*. Cambridge: Intersentia, pp. 207-251.
- Boaz, David (ed.). (1997). *The Libertarian Reader*. New York: The Free Press.
- Carlsmith, Kevin M. (2008). On Justifying Punishment: The Discrepancy between Words and Actions. *Social Justice Research* 21, 199-137.
- Caruso, Gregg D. (2012). *Free will and consciousness*. Lanham: Lexington Books.
- Caruso, Gregg D. (2018). Skepticism About Moral Responsibility. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition). Retrieved from <<https://plato.stanford.edu/archives/spr2018/entries/skepticism-moral-responsibility/>>.
- Chisholm, Roderick M. (1964). Human Freedom and the Self. *The Lindley Lecture*, 3-15. Also in Gary Watson (Ed.). (1982). *Free Will*. Oxford: Oxford University Press, 24-35.
- Curd, Martin & Cover, J. A. (Eds.), (1998). *Philosophy of Science*. New York and London: W.W. Norton & Company.
- Dennett, Daniel C. (1984). *Elbow Room. The Varieties of Free Will Worth Wanting*. Cambridge (Mass.): MIT Press.
- Dennett, Daniel C. (2003). *Freedom Evolves*, London: Penguin Books.
- Descartes, René. (1641). *Meditations Metaphysique* (1st edition). Paris: Larousse.
- Dworkin, Ronald. (2011). *Justice for Hedgehogs*. Cambridge: Harvard University Press.
- Feinberg, Joel. (1970). *Doing and Deserving*. Princeton: Princeton University Press.
- Fischer, John Martin & Ravizza, Mark (eds.). (1993). *Perspectives on Moral Responsibility*. Ithaca: Cornell University Press.

- Fletcher, George P. (2007). *The Grammar of Criminal Law. Volume One: Foundations*. Oxford: Oxford University Press.
- Frankfurt, Harry G. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy* 45, pp. 829-839.
- Frankfurt, Harry G. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68, pp. 5-20.
- Frankfurt, Harry G. (1982). What Are We Morally Responsible For. In Leigh S Cauman e.a. (eds.), *How Many Questions? Essays in Honor of Sidney Morgenbesser*, Indianapolis: Hackett.
- Gallagher, Shaun. (2012). *Phenomenology*. Basingstoke: Macmillan.
- Greene, Joshua & Cohen, Jonathan. (2004). For the Law, Neuroscience Changes Nothing and Everything. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359, pp. 1781-1785.
- Hage, Jaap. (2016). The Compatibilist Fallacy. In Jerzy Stelmach, Bartosz Brozek & Lukasz Kurek (eds.), *The Providence of Jurisprudence Naturalized*. Kraków: Copernicus Press, pp. 82-103. Also in (2017). *Revus* 32, pp. 97-118.
- Hage, Jaap. (2017). Tort Law. In Jaap Hage, Antonia Waltermann & Bram Akkermans (eds.), *Introduction to Law* (2nd edition). Cham: Springer, pp. 109-128.
- Hage, Jaap. (2018). *Foundations and Building Blocks of Law*. Den Haag: Eleven International Publishing.
- Hage, Jaap. (Present Volume). Are the Cognitive Sciences Relevant for Law.
- Hart, HLA. (1968). *Punishment and Responsibility. Essays in the Philosophy of Law*. Oxford: Clarendon Press.
- Hogdson, David. (2002). Quantum Physics, Consciousness and Free Will. In *Kane 2002*, pp. 85-110.
- Kane, Robert. (1996). *The Significance of Free Will*. Oxford: Oxford University Press.
- Kane, Robert (Ed.). (2002). *The Oxford Handbook of Free Will*. Oxford: Oxford University Press.
- Kane, Robert. (2005). *A Contemporary Introduction to Free Will*. Oxford: Oxford University Press.
- Kenneth, Jeannette. (2001). *Agency and Responsibility*. Oxford: Clarendon Press.
- Kornhuber, K. & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv für die gesamte Physiologie des Menschen und der Tiere* 284, pp. 1-17.
- Leibovitch, Adi (Present Volume). Institutional Design and the Psychology of the Trial Judge.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8, pp. 529-566.
- Lucy, William. (2007). *Philosophy of Private Law*. Oxford: Oxford University Press.

- Mackor, Anne Ruth. (2013). What Can Neuroscience Say About Responsibility? Taking the Distinction Between Theoretical and Practical Reason Seriously. In *Vincent 2013*, pp. 53-83.
- Maslen, Cei, Horgan, Terry & Daly, Helen. (2009). Mental Causation. In Helen Beebe, Christopher Hitchcock & Peter Menzies (eds.), *The Oxford Handbook of Causation*. Oxford: Oxford University Press, pp. 523-553.
- McKenna, Michael & Coates, D. Justin. (2020). Compatibilism. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition). Retrieved from <<https://plato.stanford.edu/archives/spr2020/entries/compatibilism/>>.
- McLaughlin, Brian P. (1994). Epiphenomenalism. In Samuel Guttenplan (ed.), *A Companion to the Philosophy of Mind*. Oxford: Blackwell, pp. 277-288.
- Metzinger, Thomas. (2011). The No-Self Alternative. In Shaun Gallagher (Ed.). *The Oxford Handbook of The Self*. Oxford: Oxford University Press, pp. 279-296.
- Michaels, Ralph. (2006). The Functional Method of Comparative Law. In M. Reimann & R. Zimmerman (eds.), *The Oxford Handbook of Comparative Law*. Oxford: Oxford University Press 2006, pp. 339-382.
- Miller, Alexander. (2019). Realism. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition). Retrieved from <<https://plato.stanford.edu/archives/win2019/entries/realism/>>
- Moore, Michael. (2018). "Nothing but a Pack of Neurons": The Moral Responsibility of the Human Machine. In Bebhinn Donnelly-Lazarov (Ed.). *Neurolaw and Responsibility for Action*. Cambridge: Cambridge University Press, pp. 28-70.
- Morse, Steven J. (2013). Common Criminal Law Compatibilism. In *Vincent 2013*, pp. 27-52.
- O'Connor, Timothy. (1995). Agent Causation. In Timothy O'Connor (Ed.), *Agents, Causes & Events, Essays on Indeterminism and Free Will*. Oxford: Oxford University Press, pp. 173-200.
- O'Connor, Timothy & Franklin, Christopher. (2020). Free Will. In Edward N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition). Retrieved from <<https://plato.stanford.edu/archives/spr2020/entries/freewill/>>.
- Parfit, Derek. (1984). *Reasons and Persons*. Oxford: Clarendon Press.
- Peereboom, Derk. (2001). *Living Without Free Will*. Cambridge: Cambridge University Press.
- Peereboom, Derk. (2014). *Free Will, Agency and Meaning in Life*. Oxford: Oxford University Press.
- Quine, W.V. (1976). Quantifiers and Propositional Attitudes. In WV Quine (ed.), *The Ways of Paradox and Other Essays* (Revised and Enlarged Edition). Cambridge: Harvard University Press, pp. 185-196.
- Plato. *The Republic* (many editions).

- Rawls, John. (1995). Two Concepts of Rules. *Philosophical Review* 64, pp. 3-32.
- Robb, David & Heil, John. (2019). Mental Causation. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). Retrieved from <<https://plato.stanford.edu/archives/sum2019/entries/mental-causation/>>.
- Robinson, William. (2019). Epiphenomenalism. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). Retrieved from <<https://plato.stanford.edu/archives/sum2019/entries/epiphenomenalism/>>.
- Rousseau, Jean-Jacques. (1762). *Du contrat social* (many editions). First edition: Amsterdam.
- Searle, John R. (1983). *Intentionality. An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Sinnott-Armstrong, Walter & Nadel, Lynn (Eds.). (2011). *Conscious Will and Responsibility*. Oxford: Oxford University Press.
- Sinnott-Armstrong, Walter (Ed.). (2014). *Moral Psychology, volume 4: Free Will and Moral Responsibility*. Cambridge: MIT Press.
- Smilansky, Saul. (2000). *Free Will and Illusion*. Oxford: Oxford University Press.
- Smilansky, Saul. (2002). Free Will, Fundamental Dualism, and the Centrality of Illusion. In *Kane 2002*, pp. 489-505.
- Woodruff Smith, David. (2018). Phenomenology. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition). Retrieved from <<http://plato.stanford.edu/archives/sum2018/entries/phenomenology>>.
- Smits, Jan M. (2017). *An Advanced Introduction to Private Law*. Cheltenham: Edward Elgar Publishing.
- Strawson, Peter. (1962). Freedom and Resentment. *Proceedings of the British Academy*, vol xviii, pp. 1-25.
- Strawson, Galen. (2011). The Minimal Subject. In Shaun Gallagher (ed.), *The Oxford Handbook of The Self*. Oxford: Oxford University Press, pp. 253-278.
- Talbert, Matthew. (2019). Moral Responsibility. In Edward N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition). Retrieved from <<https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/>>.
- Van Dam, Cees. (2014). *European Tort Law* (2nd ed.). Oxford: Oxford University Press.
- Van Inwagen, Peter. (1995). When is the Will Free? In Timothy O'Connor (ed.), *Agents, Causes & Events, Essays on Indeterminism and Free Will*. Oxford: Oxford University Press, pp. 219-238.
- Vincent, Nicole A. (ed.). (2013). *Neuroscience and legal Responsibility*. Oxford: Oxford University Press.
- Wegner, Daniel M. (2002). *The Illusion of Conscious Will*. Cambridge (Mass.): The MIT Press.

White, Alan R. (1985). *Grounds of Liability. An Introduction to the Philosophy of Law*. Oxford: Clarendon Press 1985.