BRAINS, MINDS, RESPONSIBILITY AND LIABILITY
Jaap Hage[1]

1. Introduction

This contribution deals with the question whether it is still possible to hold people responsible for their doings in the light of the old claim that all human behavior is determined by causal laws, and of recent findings of psychology and neurosciences suggesting that humans are less in control of themselves as may seem at first experience. This responsibility issue is important for law, because legal rules often attach a particular consequence, liability, to the pre-legal finding that somebody is responsible for an act.

Criminal liability Criminal law deals amongst others with the question under which conditions people can be punished for their behavior. One typical condition for the liability to be punished is that the suspect of a crime is blameworthy. If Joan stole a can of milk powder for baby's from the supermarket to sell it to China, her behavior was not only illegal, but also blameworthy. Under normal conditions, this blameworthiness extends from the act to the agent who performed the crime, in this example to Joan. Joan committed a crime and is blameworthy, and therefore she is liable to be punished for her crime.

Tort liability Tort law deals with situations in which one person, the tort-feasor, acted unlawfully, and some other person suffered damage as a consequence. For example, a car driver violates some traffic rule, and causes a collision with some other car, with bodily harm of the other car driver as a further consequence. In such cases the tort-feasor is liable for the damage of his victim[2]; he must provide the victim with monetary compensation of the damage. However, this liability presupposes that the tort-feasor can be blamed for what he did. If the car driver had a black-out because of the unforeseeable effects of the medicines he took, he is without blame, and then he does not have to pay damages.[3]

Blameworthiness Although there also exist substantial differences, criminal law and tort law have in common that as a starting point they presuppose blameworthiness for respectively criminal and tort liability. It is this presumption of blameworthiness which makes both criminal law and tort law vulnerable if human beings cannot, can never, under no circumstances, be blamed for what they did. And yet the presumption that human beings can sometimes be blamed for their actions has come under attack. The attack is already old. In the 18th century, the philosopher Spinoza already argued that everything, including human behavior, is determined on the basis of physical laws, and that this determination takes away every ground for blaming people. Modern neuroscience seems only to confirm the findings of Spinoza, not only by discovering the causal mechanisms that lead to human

---

[2] Well-meant attempts to pay equal attention to all sexes often lead to ugly linguistic constructions. In order to avoid this, I will follow the convention that I write as if the anonymous persons referred to in his text belong to my gender. This means that I will write 'he', rather than 'she or he'. I encourage female authors to write about 'she'.

[3] Depending on the legal system in question, this may be different if the car driver has a so-called 'strict liability'. Strict liabilities will be discussed more extensively in Section 8.2.

behavior, but also by casting doubt on the very notion of a person that acts and that can be blamed if his actions are wrong.

Responsibility and free will The attack on criminal and tort liability often assigns a central place to the notion of a free will. The argument goes by and large as follows. A person can only be liable if he can be blamed for what he did. Blameworthiness can only exist if a person is responsible for what he did and a person can only be responsible if the agent had a free will. However, the argument continues, a person's will cannot be free, because it is determined by causal laws. Therefore, a person cannot be responsible or blameworthy, and there is no ground for making him liable for punishment of for damages.

The simple seeming argument is on second sight not without complications. It is questionable whether causal determinism is true, and it is also questionable whether the existence of a free will depends on the falsity of causal determinism. Moreover, it may be argued that a free will is no necessary prerequisite for responsibility.

Overview In the following sections, we will unravel the complications step by step. First we will have a closer look at the actual practice in law of holding persons responsible and making them liable (Section 2). Then we will go into some detail concerning the argument that there is no room for responsibility because causal determinism makes the existence of a free will impossible (Sections 3 and 4). An important argument in the discussion is that the attribution of responsibility is compatible with the denial of a free will. This so-called 'compatibilism' will be the subject of the Sections 5 and 6. The existing practice of attributing responsibility is based on a particular image of man. This image has come under attack from recent findings in psychology and neurosciences. Some of these findings and their seeming implications are discussed in Section 7. At the end of this article, in Section 8, we will have a closer look at the possibility that there is no place in law for a free will and for 'ordinary' responsibility, and investigate what the implications for the law should be. The article is summarized and some conclusions are drawn in the Sections 9 and 10.

2.   Existing legal practice

Liability Liability is the typical consequence which the law attaches to responsibility. A person is criminally liable if he can be punished for having committed some crime. For example, Joe stole the Chevrolet of Bill and the judge can send him to prison for that. A person is liable according to civil law if he must compensate somebody else's damage.[4] The main reason for holding a person liable for damages is that this person through his unlawful behavior caused damage to somebody else. For instance, while playing soccer in the streets Jane (19 years old) shot the ball through the window of Harold's house. Jane is liable toward Harold for damages. This unlawful behavior is called a 'tort', and the liability in question is called 'tort liability'.

2.1   Responsibility

Criminal and civil liability have in common that in 'normal' situations they presuppose that the liable person is responsible for what he did. It is therefore important to know what responsibility is and under which circumstances a person is held responsible.

---

[4]   It is sometimes also said that a person is liable for his own damage if nobody else is obligated to compensate his damage.

Six kinds of responsibility Regrettably these questions are not so easy to answer, amongst others because the word 'responsible' is used in many different senses. Vincent (2009) distinguishes for instance between six kinds of responsibility:

1. *Virtue responsibility* is a positive characteristic of a responsible person, a person who takes his duties seriously and normally does the right thing.

2. *Role responsibility* is the responsibility that belongs to a particular role. For example, as a doctor it is your responsibility to take good care of the health of your patients. Role responsibility consists of a set of duties that pertain to a role somebody fulfills.

3. *Outcome responsibility* is the responsibility consisting in having brought about a particular result. For example, the teacher is responsible for the good results of his pupils, and the child molester is responsible for the psychological traumas of his victims. This result can be both good or bad, but most often when we speak of a person being responsible for a particular result, we are talking about a bad result. It sounds more natural to speak of the responsibility of the child molester for the psychological trauma's than of the responsibility of the teacher for his students' results.

4. *Causal responsibility* concerns the factors that made an agent do what he did. For example, the drugs he took were responsible for Harald's absent-mindedness.

5. *Capacity responsibility* concerns the mental capacities an agent has, which allow or disallow the agent to behave as a mentally healthy and adult person. Capacity responsibility would for instance be lacking in young children, or in people who are mentally ill, or under the influence of alcohol or drugs. Those people may not be responsible for what they are doing.

6. *Liability responsibility* is at stake when people use the word 'responsible' when they might as well have said 'liable'. A tort-feasor is in this sense responsible for the damage he brought about, meaning that he is liable for damages.

This set of distinctions is helpful to delineate the kind of responsibility that is at stake in discussions about liability in criminal law and in tort law. The responsibility at stake in these discussions is the responsibility of an agent for something that he did. That immediately excludes virtue responsibility, role responsibility and causal responsibility as irrelevant, because they do not concern the responsibility of an agent for his actions. Outcome responsibility is not relevant either, because the notion of responsibility that is at stake in discussions about free will and liability is not concerned with the causal connection between an act and its consequences, but with the relation between an agent and his act. Liability is something that is attached by law to the existence of responsibility in the sense we are looking for and is therefore not that kind of responsibility itself.

Capacity responsibility comes nearest to what we are looking for, but it is still not the real thing. If a person is lacking the relevant capacities, for instance the capacity to be guided by rules or the capacity to act on reasons, this is a ground not to hold this person responsible in the relevant sense. Having these capacities is not the same as being responsible, though. This becomes clear from the fact that having the capacities does not involve having acted, while the kind of responsibility we are looking for consists in a relation between an agent and an act, and therefore requires the presence of an act.

It turns out that it is quite hard to pinpoint the kind of responsibility at stake when we adduce a person's being responsible as a ground for making this person liable. We have found that this responsibility is a relation between an agent and his act. The agent must have performed the act, and there must not be reasons not to hold the agent responsible, reasons such as the lack of relevant

capacities. If such reasons are absent we assume a close relation between agent and act; the act in a sense 'belongs to' the agent. Therefore the agent deserves praise – if the act was good - or blame – if the act was bad. Moreover, - and this may explain the term 'responsible' – if the act seemed to be bad the responsible agent is required to provide reasons that justify the act. He must answer for what he did, and if he cannot he must bear the consequences in the form of blame and/or liability (Mackor 2013).

To summarize we may circumscribe the responsibility at stake in this article as follows:

'Definition' of act-responsibility

A person P is responsible for an act A, if and only if A was performed by P, and if P was not lacking certain capacities that are relevant for the ascription of responsibility. Being responsible is a moral status to which several consequences are attached. If A was good, P deserves praise for what he did. If A was prima facie bad, P must provide reasons why he did A. If these reasons are not presented, or if they are not convincing, P is liable for such consequences are blame, punishment, or damages. We will call this kind of responsibility 'act responsibility'.

Responsibility as status Notice that the above 'definition' still did not say what act responsibility[5] *is*. The definition only indicates under which circumstances the relevant notion of responsibility is applicable, and what the consequences are when this notion is applicable. This phenomenon, that it is possible to indicate when a word is applicable and what the consequences of its applicability are, without the possibility to pinpoint what the word precisely stands for, is characteristic for so-called 'status words', words that stand for a status (a meaning) that people assign to phenomena with particular characteristics, without there being a characteristic in the phenomena themselves to which the status word refers (Hage 2009).

There is no responsibility 'out there' which can be found if we look well enough. There are only characteristics of phenomena that we consider to be reasons for holding an agent responsible for some act. These characteristics typically include that the agent performed the act, and that the agent did not lack the capacities that are considered to be essential for responsibility.

2.2   Acts and agents

Attribution of acts and agency Act responsibility is a relation between an agent who is responsible and an act for which the agent is responsible. This relation is assumed to exist if, amongst others, the agent performed the act. But what does it mean that an agent performed an act? Acts and agents are not 'out there' to be discovered by an objective science; they are the result of attribution (labeling, ascription) and this attribution is based on social conventions. 'We' consider the event that the button of the mouse of Helen's computer went down under the pressure of Helen's index finger as an act, and 'we' attribute this act as clicking the mouse to Helen, whose index finger rested on the mouse button when it went down.

Acts As members of our society we share conventions concerning which events count as acts, and to whom these acts are attributed as being the agent. Perhaps these conventions are even partly the result of our genetic make-up. After all, most acts are recognized as such in most cultures of the world. However, the conventional nature of what counts as an act is very clear in the case of an

---

[5]   From now on, when we write about responsibility without additional qualification, we mean act responsibility.

internet purchase. Only a culture that has the institution of selling and buying, and that has computers and internet, will recognize a mouse click as a purchase.

The conventional nature of acts also becomes very clear in the case of acts that consist of omissions.[6] When a mother does not feed her children even though she could easily do that, this counts as an omission and consequently also as an act that can be attributed to the mother. Whether doing 'nothing' counts as an act depends on whether it was obligatory to do something. What counts as obligatory depends to a large extent on the norms that happen to exist. As a consequence it also depends on the existing norms which non-events count as omissions.

Agency Also the nature of agents depends on conventions. It only seems 'natural' that human beings are treated as the agents who performed acts. However, is it also possible that companies as such act? Is it for example possible that Facebook, the company, not its employees, infringes the privacy of its users? Or is it possible that computer, or computer programs perform acts, such as selling books on the internet? Different conventions lead to different agents.

It turns out that responsibility, acts and agency have in common that they cannot be found as such in the outside world. They all concern a status that we assign to other phenomena on the basis of conventions that we accept. See figure 1 below.
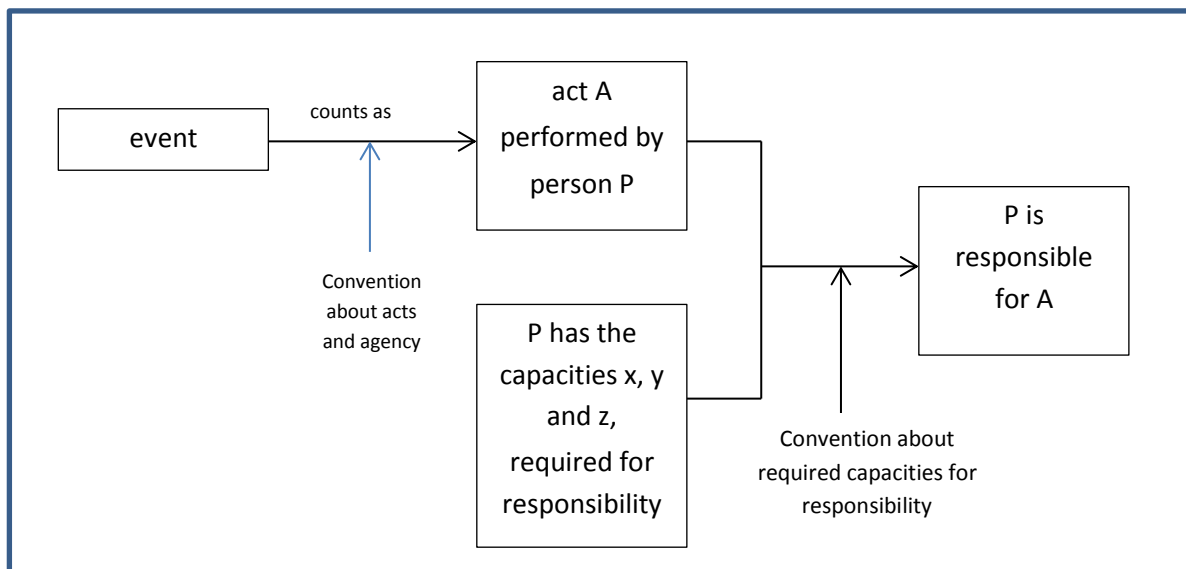

 Figure 1

Had the conventions been different, the attribution of responsibility, of agency, and of the status of an act would have been different too. An event might not have counted as an act, a person might not have counted as the agent, and the agent might not be considered responsible for what he did. And if the conventions would not even have existed, there would not have been responsibility, acts or agents, but merely physical events.[7]

---

[6] For discussion, see Keiler and Roef (2015), 60-76.

[7] We assume here, for the sake of argument, that physical events do exist 'out there', without being the result of meaning attribution. Whether this assumption is correct remains to be seen, however.

2.3 Excuses and justifications

If somebody did something wrong, that is, if some event E is classified as an act A, if A violated some norm, and if this norm violation is attributed to some agent P, then the agent P is responsible for A, *unless his behavior can be excused*. For example, Paolo drives his Fiat through the center of Rome. Suddenly a heavy explosion takes place, not very far away. Almost automatically Paolo turns his head to see what caused the loud noise, and as a consequence he drives his car into the Alfa Romeo of his predecessor, who suddenly braked, also because of the explosion. Theoretically, Paolo should have continued paying attention to the traffic, and if he had done so he would not have bumped into the other car. However, we do not blame Paolo for his instinctive reaction; Paolo's behavior was excused and Paolo is not held responsible for the accident.[8]

One way to argue for the conclusion that Paolo's act was excusable is to say that because of the explosion, Paolo at that very moment lacked the capacity to pay attention to the traffic, a capacity which he would normally have possessed.

Which capacities are required for being responsible is determined by the conventions used in society for the attribution of responsibility. The lack of these capacities can be adduced as an excuse, to argue that an agent who would normally – that is, if he would have had the usual capacities – have been responsible for an act, is not responsible in this particular case.

Insanity An important excuse in criminal law is insanity. If somebody suffers from paranoia and in fear that he will be killed hurts the presumed murderer, he can plead insanity, and under suitable circumstances[9] the insanity is counted as a valid excuse and the agent is not held responsible.

Not being responsible means in criminal law also not being criminally liable. A person who successfully pleads insanity will normally not be punished for what he did. Instead he may be forced into treatment in a mental hospital.

A somewhat 'weaker' form is the defense of diminished capacity. This defense applies if the agent's capacities were impaired by some mental disorder, but not to an extent of full legal irresponsibility. Diminished capacity, if recognized by a court, may lead to mitigation of the punishment. We will see later (in Section 8.1) that such mitigation of punishment because of diminished capacities is quite dubious if one does not believe in free will.

Justifications Excuses must be distinguished from justifications. A justification consists of facts which make that some act that normally would be wrong or illegal is under the exceptional circumstances of the justification right. For example, if a robber threatens to shoot Dennis if Dennis does not given him his money, Dennis is justified in hitting the robber on his head in order to be able to run away from the crime scene. Hitting somebody on the head is normally against the law, but the robbery justifies this normally illegal act and makes that it is lawful after all.

Justifications have in common with excuses that they take away the liability of the agent, and that is why justifications and excuses are often considered in combination, under the heading of 'defenses' against a criminal charge. In our example, Dennis would not be punishable for hitting the robber. For

---

[8]   Whether Paolo also escapes liability depends on the law of Italy; he may be held liable even though he is not considered responsible for his act (strict liability; see Section 8.2).

[9]   An example of when the circumstances are not suitable might be that the person induced the paranoia himself by taking drugs. By the way, this nicely illustrates that responsibility is not something that can be 'found' in the world, but that responsibility is attributed to an agent by the members of a community or a culture.

our present purposes, however, there is an important difference between justifications and excuses. A justification never takes the responsibility of an agent for his act away. Dennis would in our example absolutely be responsible for what he did. An excuse, on the contrary, makes that the excused agent is not responsible for his act. In a discussion of responsibility there should therefore be attention for excuses, but not for justifications.

2.4   Responsibility and liability

If an agent is responsible for his act, the law may attach liability to this responsibility. Responsibility as such is not a legal notion; it is a status assigned by social conventions.[10] The law attaches through its rules legal consequences to the presence of this pre-legal status. For example, Jane shoots a ball through the window of Harold. She is responsible for breaking the window on the basis of social conventions. The law attaches liability for damages to this responsibility. Liability is a legal status, which means that payment of the damages can be enforced by legal means. Figure 2 gives an overview of these steps and the conventions and rules involved in them.
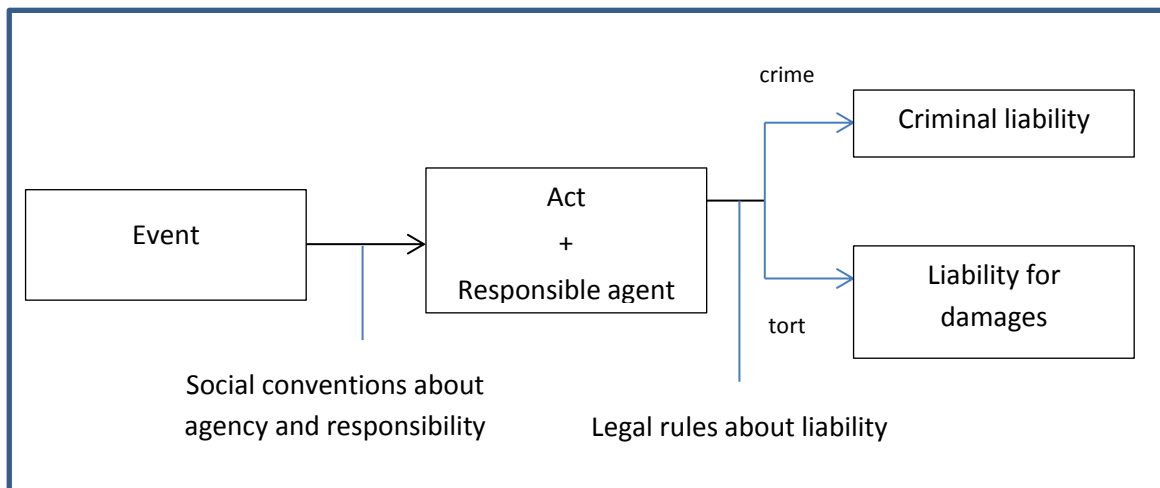

Figure 2

2.5   The image of man

The legal practice of making people liable for their actions is based on the legal rules that assign liability, but also on the social conventions that attribute agency and responsibility. It is not always made explicit, but underlying this legal practice is an image of man as a rational decision maker. The following extensive quotation from Morse (2000) gives an impression:

"In brief, the law's concept of the person is a creature who acts for reasons and is potentially able to be guided by reason. […]

The law's conception of the person as a practical reasoner is inevitable if one considers the nature of law. At base, law is a system of rules and standards expressed in language that are meant to

---

10   It is possible that the law introduces its own conventions to attribute responsibility to agents, conventions that (partially) replace the social pre-legal conventions.

guide human behavior. The law therefore presupposes that people are capable of using rules and standards as premises in the practical syllogisms that guide action. […]

The law's concept of responsibility follows from its view of the person and the nature of law itself. Unless human beings are rational creatures who can understand the applicable rules and standards, and can conform to those legal requirements through intentional action, the law would be powerless to affect human behavior. Legally responsible agents are therefore people who have the general capacity to grasp and be guided by good reason in particular legal contexts. They must be capable of rational practical reasoning. The law presumes that adults are so capable and that the same rules may be applied to all people with this capacity. The law does not presume that all people act for good reason all the time. It is sufficient for responsibility that the agent has the general capacity for rationality, even if the capacity is not exercised on a particular occasion. Indeed, it is my claim that lack of the general capacity for rationality explains precisely those cases, such as infancy or certain instances of severe mental disorder or dementia, in which the law now excuses agents or finds them not competent to perform some task.

The general capacity for rationality in a particular context is thus the primary criterion of responsibility and its absence is the primary excusing condition."

The image that is (partially) sketched in this quotation is that of a self (the person) that governs the acts that it performs through its body. This self is a rational being who, through the senses of the body, collects information about the world, and combines this information into a coherent 'image' of the world. From this image the self freely identifies reasons for and against particular acts. The self balances these reasons and decides on this basis which acts to perform. It instructs the body what to do and the body follows the 'commands' of the self. Sometimes a human being does not function in this way, and that is seen as a reason for considering him ill, injured or otherwise lacking the capacities required for responsibility. This image explains and justifies the social practice of attributing agency and responsibility and the legal practice, built on top of it, of holding agents liable for their actions.

Folk psychology and scientific psychology In the philosophical literature the image that was sketched above of mankind belongs to so-called 'folk psychology'. Folk psychology is 'the pre-scientific, common-sense conceptual framework that all normally socialized humans deploy in order to comprehend, predict, explain, and manipulate the behavior of humans and the higher animals' (Churchland 1994). It is based on how people *experience* themselves. This folk psychology should be distinguished from 'scientific psychology', which is based on *observation* and explicit *theory construction*. Some believe that scientific psychology will in the course of time replace folk psychology, just as scientific physics replaces the common-sense picture that humans have of their daily environment. Moreover, scientific psychology, which may still make use of concepts that have their origin in human experience, such as 'belief', 'desire', and 'emotion', would be replaced by a physical account of the brain/mind which uses concepts such as 'neuron', 'dopamine', and 'pre-frontal cortex' (Von Eckhardt 1994). As we will see, this opposition between the traditional way of looking at humans being as becomes manifest in folk psychology and the physical approach of the modern brain sciences plays a central role in the discussion about the existence of free will, and the sense of our social practice of holding people responsible for their actions, and the legal practice of making them liable.

2.6   Conclusion on responsibility and liability

The purpose of this section on the existing legal practice of making people liable for their actions was twofold. In the first place the section was meant to provide some background knowledge of a part of law that has become the topic of debate as result of recent developments in psychology and neurology. We have seen that normally if some event counts as an act and if this act is attributed to some agent, the agent is held responsible for the act. The term 'responsible' is mainly used if the act for which the agent is held responsible was in some way 'wrong'.

The law often attaches legal consequences to an agent being responsible for a wrong. This is called 'liability'. In criminal law this liability usually is liability for punishment. In private (tort) law, liability usually means having to pay for the damage that resulted from the wrongful act.

In the second place the section was meant to emphasize the conventional nature of classification as an act, of attribution of an act to an agent, of holding an agent responsible for his act, and of making an agent liable for the act for which he is held responsible. As we will see in Section 5, this conventional nature might salvage legal responsibility and liability from the attacks that are allegedly made on existing legal practices by insights from modern psychology and neurology.

3.   Determinism, free will and responsibility

The argument from determinism Many people argue that determinism makes responsibility impossible. Their argument goes as follows. A person can only be held responsible for acts that were the result of his free will. If determinism holds for mental facts and events, there cannot be a free will. Therefore determinism precludes responsibility. Since liability in law presupposes that the liable person is responsible for what he did, determinism also precludes the possibility of legal liability.

Overview It will be clear that if this argument is correct and if determinism holds true, the impact on law as it presently is will be enormous. The very practice of making people liable for their actions and their consequences would have to be abolished, or at least transformed considerably. It is therefore of the utmost importance for law to unravel the argument presented above and its companion assumption that determinism holds true. In this section and its subsections we will make an attempt to do so. First, in Section 3.1 we will take a closer look at the nature of causal determinism on the level of physical events. If this causal determinism is to be relevant for the existence of a free will, we need to make a transition from causal determinism to determinism of mental facts, such as the existence of a particular will, and mental events such as decisions being taken. Section 3.2 deals with the relation between the brain, as representative of physical nature, and the mind, as representative of mental phenomena. After these preliminaries follows the real argument. Section 3.3 addresses both the first step, that concludes that there cannot be a free will from the assumption that determinism applies to mental phenomena and the second step of the argument that goes from the absence of a free will to the impossibility of responsibility. Section 4 reconsiders the argument as a whole and looks forward to a seeming solution for the apparent incompatibility of determinism and responsibility.

3.1   Causal determinism?

Simply formulated, causal determinism holds that all facts and events[11] are necessitated by facts and events from the past, on the basis of causal laws. For instance, given the facts that this bar is made of iron, that it was 20 centimeters long, that it was heated during 5 minutes at a temperature of 500 degrees Centigrade, and that the air pressure was 1050 mBar (and possibly some other relevant facts), it could not have been otherwise than that the bar is now, say, 21 cm long.

Given the facts as they presently are and given the causal laws that govern physical nature, there can be only one set of facts in the near future. Since the facts of the near future similarly necessitate the facts of the somewhat later future, these latter facts are also determined by the present facts. Moreover, the present facts were necessitated by the facts that immediately preceded them. According to causal determinism, the history of the physical world is one long chain of facts that necessitate their successors in time, in accordance with physical laws.

Explanation and prediction Causal determinism is a view regarding facts and their relations over time. As such it has nothing to do with the prediction of future events or the explanation of events from the past. However, causal determinism has immediate implications for the possibility to predict events. Very theoretically, if somebody knew all the laws of nature and all the facts that existed at a certain moment, this person should be able to predict all the events in the future. He would be able to predict the length of the iron bar after its being heated, but also the result of throwing a dice. Moreover, if determinism holds we can be sure that for every event there exists an explanation on the basis of the existing physical laws and the facts of the past. There cannot be 'spontaneous' events without a cause or explanation.

Ontology and epistemology Causal determinism is a theory about facts and events and the necessary relations holding between them. As such it is an *ontological* theory[12], a theory about the nature of reality. Such an ontological theory should be distinguished from a so-called *epistemological* theory[13], which deals with our knowledge and understanding of reality. Beliefs, probability, certainty, predictions and explanation belong to the realm of knowledge, not to the realm of facts, and they are therefore dealt with by epistemology. The distinction between ontology and epistemology is important in connection with causal determinism, because the support for this view is mainly based on the success of the sciences in predicting and explaining events. For this success it is necessary to avail over 'laws' that describe regularities in nature. However, causal determinism claims more than the mere existence of regularities in nature; it claims that facts and events of the past determine in the sense of causally influence the facts and events of the future.

Is determinism true? Some people may believe that science has proven, or at least made highly plausible, that determinism is true. That is not the case. As a matter of fact, determinism is not something that can be proven true or false because it is an ontological theory, while the evidence can only relate to our knowledge of the world. Suppose for instance that there is some field, say quantum physics (Healey 2009), in which determinism seems not to hold. In that field events occur

---

[11]   Strictly speaking it is necessary to distinguish between facts and events and since causal determinism applies to both facts an events, I should properly speaking all the time write about 'facts and events'. To make the text more readable, I write instead about 'facts' or about 'events', depending on what is more suitable at that moment.

[12]   The word 'ontology' stems from the ancient Greek word 'οντωσ' that means something like 'actual' or 'real'.

[13]   The word 'epistemology' stems from the ancient Greek word 'επιστημη' that means 'knowledge'.

that could not be predicted on the basis of what went before and that appear to be completely at random. However, we should keep in mind that determinism is not a theory about predictability but a theory about necessitation. If some events necessitate some other events, for instance a metal bar being heated necessitates that the bar expands, this necessitation can be used to predict the necessitated facts from the necessitating ones. However, if it is impossible to predict particular facts, this does not show that determinism is incorrect. This impossibility to predict may be a result from a lack of knowledge of the causal law or the historical facts, rather than of the falsity of determinism.

Determinism as a paradigm Probably it is better to consider determinism to be a paradigm, a kind of preliminary assumption of the physical sciences.[14] We start doing physical science research on the assumption that all facts can be explained from other facts on the basis of physical laws, and our research largely consists in finding those laws. If we cannot find a law, we do not believe that there is no law, but only that we did not discover it yet. The unwillingness to interpret the failure to find a law as evidence that there is no law signals that we presuppose that all events have a cause, whether we already discovered it or not. Such a presupposition is a 'paradigm' of modern physical science.

As a preliminary assumption determinism should not be evaluated as true or false, but rather as fruitful or not. During the last few centuries the paradigm of determinism has proven to be very fruitful as a guideline for doing physics research, and that is a reason not to abandon it lightly. For now, it is good to assume that determinism applies to the realm of physics. However, that does not prove determinism to be true, and the problems in relation to free will might be a reason to discuss whether we should also use this paradigm in investigating mental phenomena. The outcome of such a discussion is not on beforehand clear.[15]

Applicability rather than truth Because determinism is not in the first place a thesis that is true or false, but rather a methodological assumption of scientific research that is fruitful or not in a particular scientific domain, we will write in the following not about the truth of determinism, but about its applicability. Determinism then applies to a particular domain if it is a fruitful assumption for conducting scientific research in that domain. As yet, causal determinism has proven to be a very fruitful assumption for the physical sciences such a physics itself, astronomy, chemistry, and (important parts of) biology.

3.2 Causal determinism expanded to the mind

At first sight, causal determinism only applies to physical nature and not obviously to mental phenomena, such as decisions and intentions. Many people nevertheless also apply determinism to mental phenomena, and to do so, they must make the additional assumption that mental phenomena are completely determined by physical (including chemical) phenomena.

Brain, mind and consciousness Before we continue our discussion we should make a distinction between brain and mind. Our brains are physical substances, consisting of billions of highly interconnected and interacting neural cells. Our minds are not physical entities, but are related to states of consciousness. Some mental states *are* states of consciousness, with being in pain as the obvious example. It is not possible to be in pain without experiencing the pain. Thinking about some problem is also conscious. However, our minds can also solve problems without consciously working

---

[14] The classical text about the role of paradigms in science is still Kuhn 1962.

[15] A similar discussion is also conducted with regard to the role of general laws in social science research. See the contributions by Fay and Kincaid in Martin and McIntyre 1994.

on it. It happens often that we encounter a problem, do not think about it for a while, and then suddenly we know the solution. Most likely our mind unconsciously solved the problem.

We say that our mind rather than our brain solved the problem because the solution, the knowledge how the problem is to be solved, can be made conscious. It is obvious, though, that our brain must have worked on solving the problem too, although for the brain it was strictly speaking not a problem, but rather a brain state. The problem was solved by our brain and by our mind at the same time. Moreover, in some sense there was only one process in which our brain/mind solved the problem. The question is only how our brains and our minds are related to one another.

Identity theory If causal determinism is to be applied to mental processes too, there must be a way in which the mind is 'determined' by the brain. There are at least two ways to account for this determination. One is to *identify* mental phenomena with brain states. A mental phenomenon such as the will to avoid taxes is on this view nothing else than the flip side of a particular brain state. The same thing can both be described in physical terms, as a brain state, and in mental terms as the will to avoid taxes. If the brain state as a physical state is determined by earlier physical facts and events, so is the mental state, since this mental state is, on this identity theory, identical to the brain state.[16]

Epiphenomenalism The other way to make mental states subject to determinism is to adopt epiphenomenalism. Epiphenomenalism is the view that mental states such as pain, anger, doubt, knowledge and the will to do something are merely side-effects of brain states.[17] A person with a certain brain state will also have this mental state, but the mental state does not affect the brain state. The relation between a brain state and the corresponding mental state is one-direction and comparable to[18] that between the light reflecting characteristics of an item and its color. Whether some item is red or green is completely determined by the light that this item reflects. The other way round, the color of an item has no influence whatsoever on the light that the item reflects. This color is merely an 'epiphenomenon', an added characteristic, to the reflective properties.

As epiphenomena of brain states, mental states would be determined completely by their underlying brain states. Moreover, if the brain states in their turn are completely determined by earlier physical facts and causal laws, so are the mental states. Therefore, so goes the argument, causal determinism also applies to mental states. Mental phenomena are, according to this view, completely determined by the facts of the past and could not have been different from what they actually are.

Cartesian dualism Brain and mind may seem so difficult from each other that it is hard to see them as identical. Moreover, many people find the idea behind epiphenomenalism that mental phenomena are merely a by-product of brain states unpalatable. A large part of the attraction of the identity theory and epiphenomenalism lies therefore in the unlikelihood of their traditional alternative, Cartesian dualism. René Descartes (*Cartesius* in Latin), a French mathematician and philosopher who lived around 1600 AD, held the opinion that reality consisted of two fundamentally different kinds of substance. On the one hand there is substance that exists in space, what we would now call 'physical reality', and on the other hand there is thinking substance, which we would now

---

[16] Different variants of the identity theory are discussed more elaborately in Rosenthal 1994. See also the contributions of Yoo, Baker, Van Gulick, Beckerman and Montero in McLaughlin, Beckerman and Walter 2009 and Mackor 1997, 61-77.

[17] Epiphenomenalism is discussed more elaborately in McLaughlin 1994, and in Walter 2009.

[18] Strictly speaking the relation between brains states and mental states is not merely comparable to the relation between light reflecting characteristics and color. The latter relation is a special case of the former as color is a characteristic of how the reflection of light is *experienced*.

call 'mind'. Both sensory experience and intentional action require that these two substances interact. Sensory experience requires this because signals that we receive through our eyes or ears (to mention just two of our senses) must somehow become mental in order to cause conscious knowledge. Intentional action requires interaction between the mental and the physical to allow conscious decisions to influence the movement of our limbs and mouth. If the mental and the physical are completely different substances, their interaction requires careful explanation. For instance, how can a conscious decision, a purely mental phenomenon, cause a muscle to contract? Are there causal connections, based on the laws of physics, that somehow connect mental and physical phenomena? It is hard to see how such a thing might be possible. Descartes himself assumed that the interaction occurred in the pine-apple gland, a small area of the brain. However, he did not explain how the interaction in the pine-apple gland would take place. Nowadays, Cartesian dualism does not have many adherents anymore in the circles of scientists[19], and that contributes to the popularity of epiphenomenalism and, in particular, the different variants of the identity theory.

3.3 Free will and responsibility

The argument that we are presently unraveling is that mental phenomena are causally determined by facts and events that preceded them in time and that therefore people do not have a free will, and that therefore it does not make sense to hold people responsible for their doings. Before continuing the discussion of this argument, it is useful to have a closer look at the phenomenon of free will. What precisely is the free will which according to the argument cannot co-exist with causal determination of mental phenomena?

Voluntary acts The first step in characterizing free will is to distinguish acts based on an allegedly free will from voluntary acts. An act counts as voluntary if it was not influenced by a factor that should not have influenced it. For example, if a robber threatens his victim with a gun and gives the victim the choice 'Your money or your life' the victim can in a sense choose what he will do. His will determines what he will do - at least that is what we will somewhat naively assume here[20] - but his will is influenced by the threat from the robber. Does this influence make that the victim acted involuntarily?

In this particular situation that is indeed the case, because we consider the robber's threat an inappropriate influence. However, if a shopper is seduced by the exhibited products in a candy store, his will is also influenced, but that does not make his buying the candy involuntary. The difference between the two situations is that in the first case the will of the agent was influenced by something that should not influence the will, while in the second case we have no problems with the persuasion that emanates from the exhibited candy. When our opinions about exhibiting candy change, the behavior of the customer may be counted as involuntary. Whether an act counts as voluntary is a normative issue, not a physical one. Voluntariness is not something that can be found in the 'outside world', but is a status attributed to actions on the basis of normative considerations.

Free will Voluntary acts are not the same as acts based on a free will. In both example cases, the agent decided for himself what he would do, whether he would give money to the robber and whether he would buy candy. His decisions were apparently 'up to him'. If he had decided to do something else, he would have done something else. His will seemed to be free in the sense that the

---

19 Dualism is still popular in folk psychology, though, a phenomenon that may be explained from the fact that we *experience* our mind as totally different from our body.

20 However, see Section 7.3.

act that he was to perform depended on his will, his decision to perform that act. When we want to say that a person's will is not free, it should not mean that what this person did was independent of what he wanted to do. An 'act' that was not willed is an unintentional act, but the absence of intention does not make that a free will is lacking. If somebody trips over the carpet, falls against an expensive Chinese vase and breaks the vase, he breaks the vase unintentionally. However, this is not reason to deny that this person has a free will. Having an accident is not the same as lacking free will. What should it mean, then, when we say that an agent is lacking free will?

Conventional free will There are two fundamentally different answers to this question. The one answer assumes that freedom of the will, unlike the voluntariness of an act, can be found in the 'outside world', and that the question whether human begins have a free will can in principle be answered by science. Let us call free will in this sense 'empirical free will'. We will return to this answer soon. The other answer assumes that free will is a status that we assign to agents, more or less we assign the status of voluntariness to some actions but not to others. Running ahead of our discussion of the capacities approach to responsibility in Section 6, we can already point out that on this capacities approach, an human being is attributed the possession of a free will if he has certain capacities such as thinking rationally about his actions and acting on the reasons he sincerely believes to have. The presence of a free will would on this approach be a matter of convention, and not something that can even in principle be established or refuted by scientific research. We will therefore call this kind of free will 'conventional free will'.

Empirical free will If a person is threatened by a robber it seems to be 'up to him' whether he will give his money to the robber. Empirical free will is the free will[21] that is involved in action if it is 'up to the agent' whether he will perform some act. When we say that an agent lacks free will, we mean that it is not up to this agent to decide what he will do. The agent did not, could not, decide what his will would be. His will is something that merely occurred to him, like an itch, or a sudden thought. The claim that determinism precludes the existence of a free will derives its plausibility from the thought that an agent's will is always something that occurred to the agent, and not something that *in the end* was up to the agent to determine.

Infinite regress It is not completely true that our will merely occurs to us. Sometimes our will is the result of a decision making process. For instance, when we consider different holiday destinations we gather information on the internet, ponder the different possibilities, their dates and costs, compare them to the preferences of our travel mates and ourselves, and then take a decision. From that moment on our will is directed at one particular holiday destination. Yes, that happens, but not very often, and such a decision is then the product of our will, a will that is most of the times not the result of a preceding decision making progress.

For instance, I prefer culture over nature, and that is the reason to make a city trip. After having taken this decision I want to make a city trip and this want may lead to booking one. But whence derives my preference for culture over nature? That is most likely not the result of a decision making process, and if it were, another want must have been the input to this process, and certainly this want was not the result of decision making.

Sometimes it is possible to trace back a particular will to a preceding will, which can in turn be traced back to an even more preceding will, and so on. However, this chain cannot continue forever; there cannot be an infinite regress of wills based on wills based on wills, etc. … . Somewhere there must

---

[21] In the rest of this subsection, 'free will' stands for 'empirical free will'.

have been a will that was not the product of an earlier will. This 'first will' is something that must have occurred to us like the already mentioned itch or sudden thought.

If the identity theory or epiphenomenalism is true, this 'first will' is the mental counterpart of some state of the brain. Not all brain states need to have mental counterparts, but some do. The physical processes that take place in the brain, and which mostly go unnoticed, sometimes have a counterpart in consciousness, and this counterpart might for instance be a preference for culture over nature. If causal determinism applies to processes in the brain, this brain state was necessitated by earlier brain states, perhaps some inputs from the senses, and the laws of physics. There is no role for a will in this connection, let alone for a 'free will', whatever that may be. Therefore, if the identity theory or epiphenomenalism is true and if causal determinism applies to processes in the brain, a free will is impossible.

Responsibility Suppose that Adam always wanted to buy a real Picasso. Recently he has sold his lucrative internet company to a large enterprise from Silicon Valley and now Adam can easily afford to spend 10 million euro on a painting. Adam visits an art auction and buys a real Picasso painting for 'only' 8 million euro. However, much to Adam's surprise, some of his friends blame him for spending so much money on what they call 'only oil paint on canvas'. The friends hold Adam responsible for not spending the money on charity.

Alternative possibilities Can Adam ward off responsibility by saying that he could not help what he did, that his brain made his decision for him, and that therefore his friends cannot hold him responsible? According to some theoreticians (e.g. Pereboom 2001 and Verplaetse 2011) he can, because in their opinion the very idea of responsibility presupposes a self that takes autonomous decisions. The technical term for this presupposition is that the agent has alternative possibilities.[22] Determinism excludes the existence of such alternative possibilities.[23] Even if an agent took a completely voluntary decision – and we can assume that Adam bought the Picasso completely voluntarily – he did not have alternative possibilities since the decision he actually took was necessitated by preceding facts and causal laws. The decision originated from a will that merely occurred to the agent. Such a decision cannot be said to have been taken freely, even though it was voluntary, and therefore the agent cannot be held responsible for what he did, or – perhaps better - for the act that occurred to him. Determinism makes a free will impossible and the lack of a free will excludes responsibility. Therefore, so goes the argument, causal determinism makes responsibility impossible.

4.   On four ways out

The conclusion that there is no room for responsibility, and therefore no possibility to continue our actual practice of assigning liability, is based on three premises, the second of which consists of two alternatives.[24]

---

[22]   Strictly speaking, neither Pereboom nor Verplaetse emphasizes the importance of alternative possibilities. For them the thing that matters is whether in the end the agent is the cause of his own behavior. The difference between the two seems quite theoretical, though, because if an agent is the ultimate cause of his own behavior we should expect that this agent could have chosen something else than what he actually chose. And the then alternative possibilities are back again.

[23]   Why determinism excludes this possibility is discussed in Section 6.3.

[24]   This identification of the premises on which the argument against responsibility is based was taken from Verplaetse 2011, 37. We made some modifications to Verplaetse's original list, though.

1. Causal determinism applies to the brain.

2a. Mental phenomena such as the intention to perform a certain act depend completely on the brain, either because some variant of the identity theory is true, or because epiphenomenalism is true.

2b. Mental phenomena such as the intention to perform a certain act are causally irrelevant for the bodily movements which we consider to be acts.

3. If there is no free will that causes our acts, there is no room for responsibility.

Premise 2b of this argument has not been discussed yet. Its inclusion is based on the idea that for the existence of a free will that causes acts it does not matter whether the will is determined by brain processes for which determinism holds, or whether the will is a conscious phenomenon that does not play a role in the causation of acts. In neither one of these two cases there is a free will that causes our acts: either the will is not free (2a), or it does not cause our acts (2b).

All three premises play a crucial role in the argument that leads to the conclusion that there is no room for responsibility. A proponent of responsibility therefore has a seemingly easy task: he only needs to reject one of the three premises. As a matter of fact all three premises have been disputed, and the people rejecting them have, dependent on what they reject, been labeled as libertarians (reject 1), dualists (reject 2a and 2b), zombyists (reject 2b)[25], and compatibilists (reject 3).

Theoretically it is possible to reject 1, 2a, 2b, and 3, and to be a libertarian, a dualist, a zombyist and a compatibilist all at the same time. However, most theoreticians who believe in responsibility confine themselves to rejecting only one of the premises

Libertarianism In the free will discussion, libertarians are those who deny that determinism applies to brain processes.[26] We have already seen that determinism is not so much a thesis that is true or false, but rather a paradigm, a kind of preliminary assumption that is used in scientific research. As such it should not be evaluated as true or false, but rather as fruitful or not. The problems with determinism in relation to free will might be a reason to discuss whether we should also use this paradigm when mental phenomena are at stake. However, at first sight the best way to test the usefulness of the determinism-paradigm is to see whether brain research makes good progress as long as it works on this paradigm (Lakatos 1970). As yet there seem to be no reasons why the determinist research program for brain sciences does not work.

Suppose, however, that libertarians are right, and that causal determinism is not a fruitful paradigm in brain research. That would mean that there are brain events that are, or at least appear to be, without cause. Suddenly one or more neurons 'fire' without any cause, and that leads to contraction of muscles and an event that is classified as an act by outsiders. Would the random nature of the neuron firing be a reason to hold agents responsible for the thus caused 'act'? Randomly firing neurons do not necessarily lead to a conscious phenomenon such as a will to act to begin with. But suppose that the random firing of neurons does lead to a will. Such a will would certainly be

---

[25] To my knowledge, there is no generally accepted name for those who deny that mental phenomena are causally irrelevant for acts. The name 'zombyism' was invented by myself, and will be explained later in this section. It is inspired by the idea of a zombie, an entity without a mind or, as in my present use of the term, without a mind that exercises any influence on the brain. For more on the role of 'zombies' in philosophical discussions about the mind/brain-relation, see http://plato.stanford.edu/entries/zombies/ (last visited on June 4th, 2015).

[26] There is also a libertarian ideology in political philosophy. Political libertarians assign a minimal role to the state. See Boaz 1998.

experienced as a will that merely happened to the agent. He would, for instance, suddenly feel a strong urge to buy an ice cream, completely out of the blue. If he then acts on that urge, would that be a typical exercise of free will, the ultimate reason for holding the agent responsible? The contrary seems true. If we do consider this as an act based on free will, it would certainly not be a paradigmatic example of such an act. We would even have some doubts whether to hold the agent responsible for his buying the ice cream because of a sudden urge to do so. An uncaused will is not a typical free will, and certainly not the best possible ground for holding an agent responsible.

Dualism An alternative way to rescue the will from determinism is that the will as a mental rather than a physical phenomenon does not fall under the scope of physical determinism and that in producing action, it interferes in the ordinary causal processes that go on in the brain. This is the dualist position which rejects the premise that mental phenomena such as the intention to perform a certain act depend completely on the brain.

There are two related objections against this 'solution'. One is that we have no clue how the influence of the will on the brain processes might take place. The Cartesian answer that this happens in the pine-apple gland can, without specification, not be taken seriously. And second, modern brain research provides us with no data from which we might infer that the will interferes in the 'ordinary' brain processes. On the contrary, all the available evidence from brain sciences suggests that the processes in the brain follow traditional causal pathways.

Zombyism Zombyists assume that mental phenomena *as such* [27] do not have any influence on brain processes and therefore also not on brain processes which lead to acts. They do not deny that there are mental phenomena, such as thoughts, feelings and intentions; they deny that these phenomena can influence the physical processes that go on in our brains. Since our brains cause movements of our bodies, arms, legs, and mouth, zombyists also deny that mental phenomena influence our acts. Epiphenomenalism, the view that mental states are merely side-effects of brain states and brain processes which do not feed back into our brains is perhaps the main variant of zombyism.

Most people find the idea that human beings are like zombies in that their mental states have no impact on their brains and their acts unpalatable. However, as we will see later (Section 7.3) modern neurological research does provide us with reasons that at least some conscious mental states that seem to be action guiding do not causally influence what we do.

Compatibilism The fourth way out is the adoption of compatibilism, the view that the assumption of responsibility is compatible with causal determinism and the absence of empirical free will. Because compatibilism is such an influential view, its treatment deserves a separate section.

5.   Compatibilism

5.1   Introduction of compatibilism

There is a second way, next to zombyism, to reject the premise that mental phenomena are causally relevant for acts. This second way is to assign the freedom of the will a status that is comparable to that of the voluntariness of an act: freedom of the will is not some fact (or non-fact) about the world, to be discovered by science, but a status assigned by human culture to exercises of the will. In other words, it is assumed that if free will is required for holding people responsible for their doings, the

---

[27]   The 'as such'-clause is added to account for the possibility that mental states are somehow identical to brain states (the identity theory) and that they play a causal role in their brain state-role.

required free will is not the empirical variant but the conventional variant. The idea that free will is the result of conventional assignment is a variant of the more general idea that responsibility is the result of conventional assignment. If free will is the result of conventional assignment, this free will cannot causally influence our acts.

Assignment rather than discovery Basically, the argument for compatibilism, the view that causal determinism and responsibility can go together, is quite simple. Responsibility is not something that can be discovered somewhere in the outside world. Responsibility is a status which people attribute to agents. The people who attribute responsibility are of the opinion that the agent whom they assign responsibility for an act is the one who should take the praise, or – more often – the blame for this act. Usually the reason is that they also attribute the act to this agent: he did it and therefore he is responsible for the act and often also its consequences.

Group standards The people who attribute responsibility also determine on what grounds they want to do so. The formulation of these grounds is usually not left to individual persons, but is the outcome of some unarticulated group process. The members of the group recognize responsibility when they encounter it, although there may be borderline cases. For instance we all 'know' that a mentally sane person who commits fraud by providing the Inland Revenue with false data about his income is responsible for that. We probable also agree that a mentally retarded person with an IQ below 60 is not responsible. And perhaps it is not clear whether a person who drank a little too much before making his tax statement and as a consequence forgot (really: forgot) to include an unimportant and undocumented source of income is responsible for that.

Any content The standards for determining whether somebody is responsible are developed in a social group. Theoretically, these standards can have any content. It is possible to hold an agent responsible for his own doings; it is possible to hold parents responsible for what their children did. It is possible to hold teachers responsible for what their pupils did, and to hold dog owners responsible for what their dogs did. It is also possible to hold dog breeders responsible for what dogs from their kennels did, and to hold dog breeders as a collective responsible for what any dog in the country did. And it is even possible to hold paranoid persons responsible for what they did during a psychotic attack. In short, given the 'right' standard, it is possible to hold anybody responsible for anything. All that is needed is to adopt, preferably collectively, a standard that makes the relevant persons responsible for the relevant acts. There is nothing that prevents the adoption of a standard that makes people responsible for acts that they could not influence at all, or acts that they could not help performing because they were determined to perform them. In short, given that responsibility is the result of attribution, it is compatible with determinism.[28] *Compatibilism is obviously, but also trivially, true.*

5.2   Four kinds of questions

To obtain a clear view of what compatibilism is all about, it is useful to have a look at the four kinds of question that were distinguished by Mackor (2013) regarding the issue what neurosciences can tell us about responsibility.

External questions There are two 'external questions' that address the sense of those social practices in which the notion of responsibility plays a central role. The first question asks whether (neuro)scientific research can show that those practices are fundamentally untenable because they

---

[28]   This is also a central argument in Smits 2015.

rest on false presuppositions. For example, if the very notion of responsibility presupposes the existence of empirical free will, and if empirical free will does not exist because of determinism, then, arguably, social practices than hinge upon the existence of responsibility do not make sense.

The second, more modest, question asks whether these practices may be pointless to the extent that they cannot fulfill the goal they are intended to fulfill. For instance, if we punish people to ensure that they will not commit crimes again, and if people lack the free will that is necessary to decide whether to commit crimes, the punishment would merely be an exercise of cruelty, comparable to punishing people for having blue eyes or being homosexual.

Internal questions There are also two 'internal questions' which presuppose that our social practices of holding people responsible make sense, and then ask how neurosciences can contribute to the perfection of these practices. One internal question would be whether neurosciences can teach us whether we have the best possible criteria for holding people responsible. We might, for instance, learn that people only reach the mental capacities required for responsibility at the age of forty (wisdom comes with age), and that it would therefore be better never to hold people under forty responsible for their doings.

A second internal question would be how to establish whether a person actually had the capacities that according to the existing standards are required for responsibility. Would it, for example, be possible to use a brain scan to determine whether a person can act on reasons?

It is the first of the external questions, the most radical one, which is at stake in the discussion whether causal determinism and the eventual non-existence of empirical free will preclude the possibility or sensibility of holding people responsible.

5.3   Mackor's argument

According to Mackor neuroscientific findings cannot show that the practice of holding people responsible makes no sense. In her opinion the social practice of holding people responsible if certain conditions are met embodies a standpoint with regard to the question what we *should do*. Should we praise of blame people for what they did, should we demand a justification from them or should be make them pay damages or put them in jail? Sciences, the neurosciences included can *as such* not answer the question what we should do; they can only provide us with factual information. The same holds for the paradigm of causal determinism, which is also merely factual.

However, it takes an additional step to arrive from this factual information at a judgment on what to do. For example, if science predicts that there will be an earthquake tomorrow, this prediction by itself does not suffice to draw the conclusion that we should evacuate to a safer destination. To draw this additional conclusion we must add the premise that we should avoid staying in a place where there will be an earthquake. Perhaps somebody just likes to be in an earthquake, and for this person the prediction does not constitute a reason to move. Theoretically, it might even be the case that for this peculiar person the prediction constitutes a reason to stay!

Because sciences can only give us factual information, so the argument goes, they cannot tell us what to do. And therefore they cannot tell us to abandon the practice of holding people responsible. Any argument that attempts to derive a conclusion about what we should do from the results of some science is based a logical error.[29] Science does not show us, could not show us, that the social practice of holding people responsible is wrong or should be abandoned.

---

[29]   This error has become known as the 'naturalistic fallacy'. See Moore 1903, 5 and Ridge 2014.

As the attentive reader may have noticed, this argument of Mackor's strongly resembles the observation, made in Section 5.1, that compatibilism is obviously true. At the same place it was also observed that compatibilism is *trivially* true. Compatibilism tells us that our social practice of holding people responsible for their doings is based on conventions and since conventions as such[30] are not facts, they cannot be established or denied by science.

However, conventions are not accepted for no reason. Although a convention does not follow from mere facts, facts can provide us with reasons to accept a convention. We do not hold stones responsible for the weather, and we have a reason not to do that, because we do not believe that stones can help the weather. Neither do we hold people responsible for the outcome of spinning a roulette-wheel, because we do not believe – at least normally – that people can determine the outcome of the wheel. In general, we tend not to hold people responsible for things they cannot help, and we do not only refrain from doing so but we also think that we should not hold people responsible for what they cannot help. Our social practice of holding people responsible only makes sense for those actions that people could help performing or refraining from.

Suppose now that causal determinism applies to mental phenomena and to human actions. Then, arguably people cannot help what they are doing. The actions we ascribe to people occurred to them just as the growing of their hair and toe nails. If we are consistent in our practice of attributing responsibility we should not hold people responsible for something they could not help. And if people could not help any of their 'actions', we should not hold people responsible for anything. Then the very practice of holding people responsible loses its applicability and turns out to be senseless. This is the point that authors like Pereboom and Verplaetse wanted to make in their argument against responsibility.

It seems that Mackor's argument is untenable, but there is a way to escape that conclusion, at least so it seems. This escape is based on the observation that people's capabilities, what they can and what they cannot help, is not a matter of empirical fact, but is based on convention too. On this view, no science and no paradigm can tell us which capabilities people have; it is all a matter of which capacities we ascribe to people. Although Mackor also deals with this defense of compatibilism, the presentation of the argument by Dworkin is better-known and therefore deserves to be discussed here.

5.3   Dworkin's argument

One of the more influential defenders of compatibilism was the Anglo-American legal philosopher Ronald Dworkin. In his book *Justice for hedgehogs* Dworkin provided compatibilism with an elaborate defense.[31] It is worthwhile to take a closer look at his argument.

Dworkin's argument Dworkin starts with two distinctions. The first one is between compatibilism, the view he is going to defend, and incompatibilism, the view that determinism excludes responsibility. The second distinction is between what Dworkin calls 'optimistic' and 'pessimistic

---

[30]   Although a convention itself is not a fact, the existence of some convention may be a matter of social fact, which can be established empirically. For instance, it cannot be established empirically that lying is wrong, but it can be established empirically that most people believe that lying is wrong, most of the times do not lie themselves, and want sanctions against liars. However from the fact that a convention exists as a matter of social fact, it does not follow that we should continue using it. Drawing the latter conclusion would still be a naturalistic fallacy.

[31]   Dworkin 2011, 219-252.

incompatibilism'. Optimistic compatibilists believe the determinism and responsibility cannot go together, but also that determinism is false with regard to agency. So, although determinism is incompatible with responsibility, there is no threat because determinism is incorrect for the relevant kind of facts.[32]

Pessimistic incompatibilists both hold that determinism is incompatible with responsibility, and that determinism also applies to agency. The result is that people cannot be responsible for what they did.[33] Dworkin finds that it is impossible for people to believe that pessimistic incompatibilism is incorrect. However, what people find is not decisive. Therefore Dworkin presents an argument.

Causal control Dworkin starts his argument with the assumption that we only have responsibility when we are in control of our behavior., or – in the terminology of Section 3.3 – for the things that we could help. This assumption seems to lead immediately to the conclusion that there cannot be responsibility if determinism is correct, because determinism seems to exclude control. To avoid that conclusion, Dworkin distinguishes between two kinds of control. Causal control only exists when a person's decisions are not determined by external forces in the way that determinism holds all behavior is. In other words, determinism makes causal control impossible. This means that if causal control is necessary for responsibility, determinism makes responsibility impossible.

Capacity control The other kind of control is capacity control. An agent has capacity control over his act if:

- he is conscious of facing and making a decision;
- when no one else is making that decision through and for him; and
- when he has the capacities to form true beliefs about the world and to match his decisions to his normative personality - his settled desires, ambitions and convictions.

The capacity control that Dworkin defines comes close to our actual practice of holding people responsible under normal circumstances and not holding them responsible if certain exceptional circumstances apply. What counts as normal and exceptional in this connection is answered by our social practice of holding people responsible.

An ethical choice Dworkin emphasizes, rightly, that it is not a matter of hard fact which kind of control is required for responsibility. It is an ethical issue; the question at stake is what is the best social practice for holding people responsible. Should we require causal control, or should we require capacity control? If we require causal control and if determinism is applies to the mind, we should never hold anybody responsible anymore. Our practice of holding people responsible would not make sense then. However, if we merely require capacity control, we can continue our actual practice, perhaps with some fine-tuning to take away minor inconsistencies.

So we have to choose between a practice based on causal control and a practice based on capacity control. How should we make this choice? Dworkin is very much aware of the fact that the way in which this choice is made determines which kind of control is adopted as essential for responsibility. The way we choose therefore also determines whether our present practice of holding people responsible under certain circumstances makes sense.

---

[32] This view was discussed in Section 4 under the heading 'Libertarianism'.

[33] This view will be discussed in Section 8 as 'hard incompatibilism'.

It is therefore somewhat surprising that Dworkin writes that we should make this choice by finding the *best possible interpretation of our actual practice*. According to Dworkin we should start from our present practice, try to find its underlying ideas, including its underlying image of man even though Dworkin does not mention that explicitly. From that starting point we should try to determine which kind of control best fits our actual practice. It should not come as a surprise that capacity control best fits with our actual practice, because capacity control was *defined* as the kind of control that is required by our actual practice of assigning responsibility.

Naturalistic fallacy From a logical perspective, the argument presented by Dworkin is an instance of the natural fallacy, the fallacious derivation that something ought to be the case from the fact that it is actually the case. Dworkin's argument, when all the elaborations are stripped away, boils down to it that we should choose for capacity control for our practice of assigning responsibility, because that choice fits best with our actual practice. We do it this way and therefore we should do it this way. That Dworkin's argument consists of a naturalistic fallacy does not mean that his conclusion is false; it only means that the argument that Dworkin offered for the continuation of our actual practice of assigning responsibility does not support its conclusion. It only convinces those who were already convinced to begin with.

5.4   A counter-example

The weakness of Dworkin's argument becomes more clear when we take a look at a similar argument about a practice which most of us would not support: drawing cards to predict the future.

Suppose there exists a community in which the practice has arisen to predict the quality of an upcoming marriage by drawing playing cards. The prospective groom drinks a 'predictive potion', a magic formula is pronounced, and then the groom draws, one by one, at most five playing cards from a shuffled deck. The rules are that if from the five cards three or more are red, the marriage will be happy, and otherwise not. However, if the very first card happens to be the Ace of Spades, the marriage will be happy anyhow, and the drawing of cards is not continued.

Interpretation Suppose that this practice has existed for some time, when unexpectedly a 'hard case' arises. The first card drawn by the groom is the Ace of Hearts, and the second card is the Ace of Spades. On one interpretation of the rules of the practice, the groom should continue the drawing until he has five cards. Some, however, favor a different interpretation. The Ace of Hearts is the most important red card, and as such has clearly predictive power, they say, for a happy marriage. And then the second card is the Ace of Spades, which should have predicted a happy marriage when it would have been the first card! Such a combination surely indicates that the marriage will be happy, and continuation of the card drawing procedure would be useless.

Understanding from within? Which side is right? If this practice of card drawing is comparable to the law as Dworkin sees it, we should try to understand the practice from within. Why do people believe that red cards predict a happy marriage (ask them!) and why do they assign a special role to a single black card, the Ace of Spades, when it is drawn as the first card? We should try to find the best possible interpretation of the actual practice and then use this interpretation to determine which side is right in the dispute about the hard case. What we should NOT do according to Dworkin is to step outside the practice and to ask whether the very practice of card drawing to predict the quality of the marriage makes sense. We work within a practice and to determine what is the best way to deal with a hard case arising from the practice, we should interpret the practice.

External criticism Not many would agree that in the case of this example we should take the practice as a whole for granted and only argue from within the practice to find the best solution for the hard case. Most would say that drawing cards to predict the quality of marriages does not make any sense and that arguing from the presumption that it does, is misguided. The proper way to deal with the hard case is to use it as an occasion to stop doing what was nonsensical all the time!

In a similar way we should ask whether the very practice of holding people responsible makes sense, and we should not answer that question by merely looking at the practice as it actually is and by giving the practice its best possible interpretation. The practice of holding people responsible should be evaluated in the light of *all* available knowledge. If that knowledge includes the applicability of determinism to mental phenomena – an issue that is not at all settled as we saw in Section 3.2 - then determinism should play a role in judging our actual practice of holding people responsible. Then we might use the notion of causal control to determine whether a person is responsible for what he did, and the outcome might be that nobody is ever responsible for any of his doings, and that the very practice of holding people responsible makes no sense. A hard case, for example a case about somebody who only seems accountable to a diminished degree, should not be seen as an occasion to interpret our present practice, but as an occasion to realize that what we were doing did not make sense from the very start.

6. Capacities

6.1 The capacities approach to responsibility

Dworkin's argument for the capacity approach to responsibility may be fallacious, but that does not mean that the capacities approach is wrong. We should therefore independently investigate what its virtues are. The underlying assumption of the capacities approach is formulated well in the following quotation from the work of Morse[34], which is repeated below:

> "At base, law is a system of rules and standards expressed in language that are meant to guide human behavior. The law therefore presupposes that people are capable of using rules and standards as premises in the practical syllogisms that guide action. […]

> Unless human beings are rational creatures who can understand the applicable rules and standards, and can conform to those legal requirements through intentional action, the law would be powerless to affect human behavior. Legally responsible agents are therefore people who have the general capacity to grasp and be guided by good reason in particular legal contexts. They must be capable of rational practical reasoning. The law presumes that adults are so capable and that the same rules may be applied to all people with this capacity."

Which capacities? It is assumed that humans in general have the capacity to use rules to guide their action. This capacity is general, shared by most adult humans, and therefore human beings can generally be held responsible for their doings. However, sometimes there are special circumstances which make that an agent lacks this capacity to have his conduct guided by legal rules. When such circumstances are present, this may be a reason not to hold a human agent responsible for his acts. The test for responsibility if a legal rule was violated is, allegedly, therefore whether in the concrete

---

[34] Morse, o.c.

case there were special circumstances that took away the agent's general capacity to be guided by the relevant rule.

The same point was made more concrete by Dworkin[35] who assumed that an agent has capacity control over his act if he is conscious of facing and making a decision, when no one else is making that decision through and for him, and when he has the capacities to form true beliefs about the world and to match his decisions to his settled desires, ambitions and convictions.

Capacities and compatibilism The capacities approach is used to defend compatibilism, the view that our practice of assigning responsibilities is compatible with determinism. At first sight the compatibility of the capacity approach and determinism is obvious. According to determinism a human agent who violated a rule could not have had the capacity the obey the rule. All behavior was necessitated by causal laws and the preceding facts and therefore the rule violation was also necessitated. There could not have been a capacity not to violate the rule. Since the human agent apparently lacked the capacity to comply with the rule, he should not be held responsible. Therefore the capacities approach and determinism lead to the same conclusion: nobody should ever be held responsible for his doings.

Clearly this is not what adherents of the capacities approach have in mind. They assume that our present practice of holding most human agents responsible for most of their acts, is right. To do so consistently, they must also assume that most human beings who violated rules in particular circumstances had the capacity to comply with these rules *under those circumstances.* Such an assumption seems incompatible with determinism, and therefore the question needs to be addressed how compatibilists can assume that the actual practice of assigning responsibilities can go together with determinism. To that purpose we must delve a little deeper into the nature of capacities and possibilities.

6.2 What is a capacity?

An agent has the capacity to do something if he can do it. But what does that mean? If Louise actually undersigned her exam because the rules required that, it is obvious that Louise could undersign her exam. More in general, if an agent performed some act, he had the capacity to do so. However, we are more interested in capacities in cases where somebody did not do what he had the capacity to do. If Louise violated the exam rules and did not undersign her exam, how can it be established whether she had the capacity to undersign?

Possible worlds theory Capacities, and more in general, possibilities are the most interesting in case they were not realized. It is notoriously difficult, however, to establish the existence of possibilities, including capacities, in case they were not realized. To deal with this problem, a thinking device was constructed: possible worlds theory.[36]

Necessity The basic idea underlying possible worlds theory is that something is necessary when it is the case whatever else may be the case. For instance, whatever the other facts may be, in any case every colored object has a surface and whatever the other facts may be, the number 5 is bigger than the number 3. Therefore necessarily every colored object has a surface and necessarily 5 is bigger than 3.

---

[35] Dworkin, o.c.

[36] The idea of possible worlds theory can be traced back to the German Philosopher Leibniz (1646-1716), who defined necessity as that what is the case in all possible worlds.

A different way of expressing that something is the case whatever else may be the case is to say that it is the case *in all possible worlds*. In all possible worlds every colored object has a surface and in all possible worlds the number 5 is bigger than the number 3.

Possibility The real world consists of all the facts as they actually are, while a different possible world contains a set of all facts as they might have been under different circumstances. Actually - in the real world – John has brown hair, but under different circumstances, in some other possible world, John is red-headed. Because there is some alternative possible world in which John is red-headed, it is possible that John is red-headed. In fact, he is not, but he might have been. Something is possible if it is the case in some possible world. That may be the actual world, but that is not necessary.

Capacity In the actual world, Louise undersigned her exam, but in some other possible world she did not. Therefore, actually Louise undersigned, but it would have been possible that she did not undersign. This captures the notion of a capacity quite well. *We might say that an agent has the capacity to do something if there is a possible world in which the agent does it.* That would mean that Louise has the capacity to undersign her exam if there is some possible world in which she undersigned her exam.

6.3   Possible worlds and constraints

We now have a definition of what it means that a person has a certain capacity, but it may seem that this definition has replaced one problem – the nature of capacity – with another problem, the nature of a possible world. What makes a set of facts a possible world?

Logical necessity Here the notion of a constraint plays a role. Not all sets of facts can go together. To give an obvious example: the fact that it is raining (here and now) cannot go together with the fact that it is not raining. Incompatible facts cannot be part of one and the same possible world. That is a constraint on possible worlds. A logical constraint in this case, because it is a matter of logic that incompatible facts cannot go together.

Physical necessity Next to logical constraints, there can also be physical constraints. The laws of physics can be interpreted as constraints on worlds that are physically possible. It is, for instance, physically possible that a metal bar is red, but it is physically impossible that a metal bar is heated but does not expand. There is no physically possible world, no world that satisfies all the physical constraints, in which a metal bar is heated but does not expand. And neither is there a physically possible world in which something travels faster than light in vacuum.[37]

Relativity of necessity and possibility We are now in a position to define possible worlds more precisely. A possible world is a world that satisfies a set of constraints. A logically possible world satisfies the laws of logic; a physically possible world satisfies the laws of physics. A particular world counts as possible if it satisfies a set of constraints. The nature of this possibility depends on the nature of the constraints. Logical constraints define logically possible worlds and logical possibility; physical constraints define physically possible worlds and physical possibility. Only relative to a set of

---

[37] Obviously these examples of physical possibility work with generally available knowledge of physical laws. This knowledge may turn out to be false, and then our ideas about what is physically necessary or possible turn out to be false too. This goes to show that necessity and certainty are not the same things. Something may be uncertain, but if it is true, it is also necessarily true. For more elaborate discussions of the relation between necessity and certainty, see Kripke 1972.

constraints does it make sense to ask whether something is possible or necessary. Necessity or possibility *tout court*, without being made relative, does not make sense. Every time when somebody claims that something is possible, it is legitimate to ask relative to which set of constraints it is possible. If the set of constraints cannot be specified, the claim about possibility is too obscure to make sense.

Facts as constraints Both logically and physically it is possible that John is red-headed, but is it still possible if we take into consideration that John just finished dying his hair brown? That is apparently not the case, and it is worthwhile to consider more closely why that is not the case.

Both with logical and with physical necessity (and possibility) the necessity is the result of constraints that consist of laws, the laws of logic and of physics respectively. A law expresses a necessary general connection between types of facts, for instance the type of fact that something is a metal bar that is being heated and the type of fact that this something expands. When we speak of possible worlds, such laws are the most obvious constraints to take into account. However, it Is not necessary to take only laws into account as constraints. There is no fundamental reason why particular facts should not be considered as constraints too. One such a fact might be that John just finished dying his hair brown. Given that fact, it is necessarily the case that John's hair is brown, and impossible that his hair is red. And given the fact that the train Dobrochna was on departed five minutes ago, it is impossible that she was seen at the railway station one minute ago.

In particular in connection with the claims of determinism it is important not to take only laws into account as constraints on possible worlds, but also facts. If it is claimed that Louise could not help submitting the exam without having undersigned it, this claim will probably not only be based on the laws of nature (purely physical necessity), but also on facts concerning Louise's personal history.

6.4   The relativity of capacity

An agent has the capacity to do something if there is a possible world in which the agent does it. Now we know that this specification of capacity is still too vague: we also need to specify relative to which set of constraints the capacity exists. The crucial question is: which set of constraints should be taken into account in determining whether a particular agent had the capacity to perform some act, or to refrain from performing it.

It is clear that in determining the capacities of a particular agent, we should take some personal characteristics of this agent into account. Going only by the laws of physics which are the same for everybody, every agent would have the same capacities. That would be an unattractive finding, and to avoid it, we must take personal characteristics into account in determining which capacities some agent has. But which personal characteristics should be taken into account? If the agent cannot write, we should most likely take that into account. So if Louise could not write, she did not have the capacity to undersign her exam and most likely she should not be held responsible for not undersigning it.[38] Should we also take into account that the agent was strongly motivated to violate a norm? To return to an example that I used before, suppose that a kidnapper held Louise's baby and required from Louise that she would not undersign the exam. Almost paralyzed by fear that something would happen to her baby, Louise does not undersign. Did she have the capacity to sign? Would that have been different if Louise was a drug addict who could only score if she did not

---

[38]   That might be different if it was Louise's fault that she cannot write, but for now I will ignore the possibility of responsibility without capacity.

undersign the exam? If we want to distinguish between the latter two cases, would that be a distinction based on a moral judgment regarding what *ought to* motivate Louise?

Collapse of the distinction between what is possible and what is actual Stepping back from this casuistry, the general issue raised by determinism is the following: if all facts regarding to an agent are taken into account, as well as all physical laws, the only thing that an agent could do is what he actually did. The distinction between what an agent did and what he had the capacity to do makes only sense if not all facts are taken into account as constraints on what is possible.

Capacity as a normative issue Then the question arises which facts should be taken into account, and which facts should not. Capacity becomes a normative issue, the issue which acts *should* be left out of consideration to determine what else the agent could have done next to what he actually did. Perhaps this seems an acceptable approach; after all it is what lawyers are actually doing when they ask whether a criminal suspect could have acted differently than he actually did. We should realize, however, that if we make capacity a normative notion, we cannot anymore adduce the capacity of an agent as a reason for holding the agent responsible. What we actually do is to give one single normative judgment concerning both the capacities and the responsibility of the agent. Either we judge the agent to have the relevant capacities and to be responsible, or we judge him to lack the capacities and not to be responsible. This judgment cannot be founded in the capacities of the agent, because these capacities are themselves part of the judgment. To do otherwise would be a circular argument.

Attribution of capacity, free will and responsibility The last observation that there may be a single judgement, covering both the presence of a capacity to have acted differently and the assignment of responsibility for the act that was actually performed, touches the core of the compatibilist approach. The argument from causal determinism to the conclusion that there can be no responsibility in a sense works from the 'bottom' upward: everything is causally determined, therefore an agent does not have a free will and therefore the agent could not have acted differently and therefore the agent cannot be held responsible. The 'bottom' of this argument is taken to express a hard fact about the world we live in and the rest follows from this hard fact.

Compatibilists work in a fundamentally different way. Responsibility is something we attribute to agents, and in doing so we attribute the status of an act to an event that took place, the status of agent to the person who was causally involved in bringing about this event, we attribute free will to the agent and with free will also the capacity to have acted differently. These are in the compatibilist view not different argument steps that build upon each other, but one act of assigning meaning to ourselves and to the world that surrounds us. This meaning encompasses acts and agency, free will and capacities, responsibility and liability.

6.5   Preliminary conclusion concerning the capacities approach

We saw that Dworkin's argument for the capacities approach to responsibility rested on a naturalistic fallacy: this is how it is done and therefore this is how it should be done. In this section we took a closer look at the capacities approach to see whether it is attractive even if we ignore Dworkin's argument. Central to the capacities approach is the assumption that human beings are normally capable to comply with the rules of law and that therefore they should normally be responsible for norm violations. However, there may be special circumstances in which the agent lacks the capacity

to comply with the applicable rules, and that would be a reason not to hold the agent responsible for eventual violations.

The central question is what it means that an agent lacked the capacity to comply with a norm. If the capacities approach is to lead to different results than the approach, inspired by determinism, not to hold anybody responsible under any circumstances, it must assume that sometimes an agent violated a norm even though he had the capacity to comply. We saw that this assumption would only cut ice if, in determining the capacities of an agent, not all facts about the agent are treated as constraints on what counts as possible. Some facts should be left out of consideration to allow the agent the 'freedom' to choose between norm compliance and norm violation.

The problem here is that there are no obvious criteria to determine which facts should, and which facts should not be treated as constraints on what the agent could do. If the choice which facts are treated as constraints is the outcome of normative decision making, it is not possible anymore to adduce the capacities of the agent as a reason for holding the agent responsible. Doing this anyway would amount to a circular argument along the following line: we want to hold the agent responsible for what he did and therefore we do not treat the facts that caused him to violate the norm as constraints that define the agent's capacities.

For now we may conclude that the capacities approach to holding agents responsible is the outcome of normative decision making without foundation in an independent notion of capacity. The argument based on determinism that our practice of holding people responsible does not make sense therefore applies equally to the capacities approach as to the traditional image of man as rational decision maker. This should not come as a surprise, since the capacities approach is based on this traditional image of man.

All of the argument above was based on purely philosophical analysis. No findings of modern (neuro)science were necessary to draw the conclusions that we reached above. Some might see this as a strength of the above argument: the argument does not depend on the most recent findings of science which might after some time still turn out to be false. Others may have more faith in robust empirical science than in philosophical speculation, however. For the latter, we will explore some recent scientific findings to see what modern (neuro)science has to say about compatibilism.

7. Cracks in the traditional image of man

7.1 The traditional theory of mind

Mentally healthy people tend to consider themselves as persons with one single mind. This mind is experienced as continuous in time, although there are interruptions in our consciousness, for instance when we sleep or are under anesthesia. When everything goes well a person's mind seems to be in control over his acts. An agent consciously decides what to do - an operation of his mind – and this decision causes, or at least causally contributes to, the bodily movements by means of which he executes his decision. In taking this decision, the agent can balance all the reasons for and against he thinks of. He knows about the existence of these reasons as the result of sensory perception, provided by the senses that are part of his body. All the information from his senses is combined into an 'image' of the world around him, and this image informs the agent what reasons he has to act in a particular way. This image, which mentally healthy people tend to have of themselves and of others, may be called the traditional image of man.

The self According to this traditional image, a person is characterized by a single self, which fulfills at least two functions. The first function is that the self collects all the information that somebody receives via his senses, either directly (feeling or smelling), or indirectly (reading, and hearing a lecture). The self not only collects this information, but also processes it, for instance by removing inconsistencies, and stores the results in long term memory. The self is the end point of the information stream that goes into the person.

The second function is that the self initiates action. It does so by sending 'commands' to an agent's muscles, but only after having decided what to do. Such a decision is often taken consciously, and then it may be the result of explicit balancing of reasons for and against. The decision may also be taken unconsciously. In the latter case, the self has no conscious access to the decision making process, but that there must have been such a process is clear from the fact that the muscles of the agent did receive a 'command' to contract or relax. The self is not only the end point of the incoming information stream, but also the starting point of outgoing action.

Cracks in the traditional image Give and take a few details, the traditional image describes how most of us experience their own mental lives and personalities. A person IS his perceiving, deliberating, deciding, and body-steering mind. At first experience, this seems to be unproblematic, but at second thought there are cracks in this simple image we have of our selves. Part of these cracks are provided by experiences from daily life with which we all are acquainted, but which we usually ignore if we – wise enough - do not think too much about ourselves. Another part is provided by relatively recent psychological and neurological research.

7.2 Examples from daily life

At first impression the traditional theory of mind describes quite well how we experience our own minds. However, if we observe our mental behavior somewhat closer, there are many signs that this traditional image is not completely correct.

Unconscious behavior Behavior that we perform unconsciously, or at least less consciously, is one case in point. When we travel to a destination where we go quite often, for instance work or school, it happens that we arrive at our destination without having experienced the journey consciously. We also have no, or few, memories of what happened on the way, for instance of the traffic lights for which we stopped. And yet, while making the journey we must have taken many decisions, for instance to stop for the travel lights. Without such decisions we would have been involved in many accidents. The phenomenon of sleep walking has some similarities with this form of unconscious acting. While being asleep, we can apparently do many complicated things, including committing a murder.[39] These examples illustrate that much of our behavior is not the result of conscious decision making, and not based on information that is processed in the 'normal' way, that is by passing through our consciousness and leaving traces in our memories.

Conflicting processes Another type of examples concerns an apparent battle for the use of our mouth and limbs. Slips of the tongue are a well-known phenomenon that belongs to this category. People sometimes say things that they 'really' did not want to say, or they mix two sentences that they wanted to pronounce both. Phenomena like these might be explained from the fact that two (or more) mental processes both claim the use of our mouth and that this battle is not neatly won by

---

[39] See http://en.wikipedia.org/wiki/Homicidal_sleepwalking (last visited 30/5/2015) for an overview of allegedly historic cases. A vivid description of one of these cases can be found in Lamme 2013, 7-10.

one of the contestants before the mouth started to speak[40], which would be the normal course of events, but that the contestants intermittently grasp control while the mouth continues speaking.

Other examples of an apparent battle for control between different mental processes are when a person starts to walk to one destination, even though 'he' decided one second before to go to some other destination, or when he wants to put a plate in a cupboard, but inadvertently opens the refrigerator, apparently to put the plate there.[41] There are even stories known of people whose one arm tries to undo what the other arm accomplished.[42]

Concentration A third category of examples concerns problems with concentrating. Everybody who has ever fallen in love will recognize the phenomenon described in the song title 'All kinds of everything (remind me of you)'.[43] For a person who was engaged in an emotional conversation, it is difficult to focus on something else. It is as if the mental process recapitulating the conversation and continuing it in soliloquy dominates the focus of this person's attention even though 'he' wants to focus on driving his car safely. Then the question may be asked who this person really is, the mental process that continues the conversation, or the car driver. Or is it still a different mental process, consciousness or attention, for which the two other process fight?

It is not easy to provide questions like these with satisfactory answers. However, the examples – and there are many more[44] - do illustrate that there are reasons to cast doubt on the traditional image of a central self who decides what to do and then executes the results of these decisions by means of its body's mouth and limbs.

7.3 Rational agency

Explanatory reasons One aspect of the traditional image of man is that there are a least a number of decisions that are taken for a reason. This formulation, 'taken for a reason', is deceptively simple. The word 'reason' has at least two meanings. One meaning is closely related to 'cause'. This meaning is relevant in the sentence 'The reason that he screamed was that somebody hit his thumb with a hammer'. It usually does not make much sense to scream when being hit, but being hit nevertheless explains the screaming. It is also possible to predict that somebody will scream when he will be hit with a hammer on his thumb. It has become customary in the literature about reasons[45] to call reasons in this sense 'explanatory reasons'.

Guiding reasons The other meaning of 'reason' has to do with the rationality of an act. A person can have reasons for or against doing something; these reasons make the act the rational, respectively an irrational thing to do. For instance, the fact that it is cold inside is a reason why Giovanni should put a coat on. Perhaps Giovanni is stubborn and is not at all motivated by the cold to put a coat on, but he nevertheless has a reason to do that. Reasons in this sense are called 'guiding reasons'.[46]

---

[40] Notice the move in formulation from an acting person to an acting body part.

[41] These examples were drawn from personal experiences of the author, but are hopefully also recognizable for others.

[42] See Brian and Chatterjee (2004), referenced after Lamme 2013.

[43] See https://www.youtube.com/watch?v=vpKypJkzrkI

[44] Lamme 2013 abounds with examples.

[45] For instance, Raz 1975, 15-19.

[46] Raz *ibidem*.

Dual role of reasons If somebody acts for a reason, both senses of 'reason' play a role. If Giovanni put his coat on, he had a guiding reason to do so. Moreover, the fact that it is cold has, let us assume, motivated Giovanni to put his coat on. The guiding reason was therefore also an explanatory reason. When human beings act, this coincidence of explanatory and guiding reasons occurs quite often, and may even be said to be the normal situation (Davidson 1963). People tend to act rationally, and therefore guiding reasons (which make an act rational) will normally also be explanatory reasons (which brought the act about).

Rational agency This idea plays a central role in the traditional image of man: being rational, a human agent is guided by the reasons that tell him what is the best thing to do and in this way the guiding reasons also explain the agent's doings. However, the traditional image adds another element to this account. This other element is that it is the awareness of the presence of a guiding reason, a state of consciousness, which plays a causal role in the generation of the act. Spelled out for the example about Giovanni, this leads to the following result: It is cold outside and Giovanni is conscious of this fact. This state of consciousness causally contributes to Giovanni's brain sending an impulse to the relevant muscles, making Giovanni put his coat on. In this way, the guiding reason, the fact that it is cold, causes Giovanni's behavior via the intermediary of a mental state: Giovanni being conscious of the fact that it is cold.

Libet's experiment Recent neuroscience provides us with reasons to assume that this traditional image may be wrong, to assume that an agent's consciousness may not always play a causal role in the generation of this person's acts. The central findings in this connection stem from research by Libet (Libet ea. 1983). Libet showed his test subjects a button with behind it a running clock. He asked his subjects to push the button anytime they liked, but to memorize the position of the clock hand the first moment they felt the urge to push the button. It turned out that the urge to push the button typically was felt 0.3 seconds before the button was actually pushed. Apparently the actual act (pushing the button) was the result of the earlier decision to do so.

However, Libet also registered the brain activities of his test persons. As was to be expected the act of pushing the button was preceded by activity in the motor cortex of the brain, approximately 0.1 second before the muscles of the test subjects actually contracted. However, there was a relatively lengthy preparatory brain activity that consisted of building up a so-called readiness-potential (*Bereitschaftspotential*). This process started about a full second before the test persons actually pushed the buttons. This is more than half a second before the test subjects consciously decided to push the button. Apparently the brain 'took the decision' to push the button more than half a second before the person who actually performed the act was aware of it.[47] It seems to follow that it was not the conscious decision that initiated the process that lead to the button pushing, but an unconscious brain process.

Libet's experiments (there was more than one experiment) have been the object of much discussion, but the least that can be said is that the traditional image of rational decision making, with a central

---

[47] It is very tempting to stick to mental terminology such as 'taking a decision', even if unconscious brain processes are discussed. We should keep in mind, though, that the very point of the present discussion is that decision making in the traditional sense of the expression, according to which a decision initiates an act, is misguided. In that traditional sense of decision making, acts are not brought about by decisions. There is no decision leading to an act, only an unconscious brain process.

role for consciousness, has become controversial.[48] It seems that it is not always (or always not?) the conscious decision to perform an act that causes the performance of the act.

Fabulation However, there is more than only the finding that actions are not (always) initiated by conscious decisions. There is also reason to assume that agents do not know what made them perform the actions they actually performed. Notice that we do not talk of actions performed without thinking, or in a state of diminished consciousness; we are talking of fully conscious taken decisions resulting in an act conform the decision. There are different kinds of evidence that support this claim (Lamme 2013, 120-147, 181-211). One experiment, which is here only described incompletely[49], concerns test persons who were asked which of three glasses of Cola tasted best. The glasses came from bottles clearly labeled as Coca Cola, Pepsi Cola and Shoppers Cola (a cheaper brand). As a matter of fact, all three glasses contained Shoppers Cola. In a test group, 80% of the persons declared that the 'Coca Cola' tasted best, and they gave plausible sounding reasons for this preference, such as that the 'Coca Cola' glass had more bubbles, or a better (sugar) taste. Actually, the persons who set up the experiment has manipulated the members of the test group to make them prefer the 'Coca Cola' over the 'Pepsi Cola'.

7.4 Conclusion

The experiment by the research group of Libet and the cola experiment provide evidence for the theses that not all human actions are the result of conscious decision making and that where they seem to be the result of it, agents do not always know what caused their conscious decisions. Both findings undermine the traditional image of a self that consciously decides what to do on the balance of reasons, and that steers a person's actions. To the extent that this traditional image underlies our social practices of attributing actions, capabilities and responsibility to agents, the psychological experiments and the findings of neuroscience at the same time also undermine these social practices.

The examples of seemingly competing brain processes described in Section 7.2 have a similar effect. Suppose that a person has a single body, but that his brain runs several, perhaps even many, different processes of which at most one is conscious. These processes do not necessarily exchange information, and may in turn grasp control over the person's mouth or limbs, thereby making him act. If the law holds somebody responsible for an act, it is a single person who is held responsible. However, if the above theory of competing brain processes is by and large correct, the only thing that is 'single' is the body.[50] Should we therefore say that we hold some *body* responsible for acts performed by one of the brain processes[51] running in this body? Should we only hold the process that initiated the act responsible - meaning that something else than a person is held responsible - and let this process compensate the damage or spend time in jail? It is not even imaginable how a brain process *as such* could spend time in jail. Or should we lump all the brain processes together,

---

[48] This is the central theme of Sie and Wouters 2010.

[49] The description of the experiment was taken from Lamme 2013, 133-135. The full story, which is a bit long to describe here, is even more convincing concerning the message that people do not know what determines their choices.

[50] Even the assumption that there is a single body can only be upheld because we 'lump together' the different body parts to make a single body.

[51] Notice how it is both 'natural' and weird to attribute acts to brains processes. Somehow we need an agent, because we see an act, and if the agent is not the person, then it must be the brain process. And yet, brain processes are not the typical entities to which we ascribe agency.

call their collection a person, and hold this collective under the name 'person' responsible for all acts initiated by one of the members of the collective? Even then it is hard to imagine how this should be accomplished.

These questions may sound stupid, but they impose themselves on anybody who takes the competing processes theory of mind serious. However, perhaps there is no need to answer these questions if we totally abandon the desire to hold persons responsible for acts. Implicit in the idea of responsibility is often an element of blame. Does this very idea of blame still make sense if there is no central control over a person's acts, if there are only multiple brain processes. Should we really blame brain processes and their mental counterparts, or blame the collective of which they are parts? It seems not.[52] Apparently the idea of the vanishing self deals an even more devastating blow to the idea that humans are responsible for their acts than the idea that all events, and therefore all human acts, are determined by earlier facts and causal laws.

8. Incompatibilism

Suppose that determinism or some variant of the competing brain processes theory of mind is correct and that, amongst others for that reason, the conventions that attribute responsibility to human agents do not make sense. We can call this view hard incompatibilism. (We will define soft incompatibilism later.) According to hard incompatibilism, it does not make sense to hold people responsible for "what they did". What should be the implications of hard incompatibilism for law? There are two fields of law in which responsibility plays a crucial role. They are the fields of criminal law and of liability law.

8.1 Hard incompatibilism and criminal liability

One principal in criminal law is that a person can only be punished if three main conditions are met:

1. There was an illegal act.
2. This act can be attributed to an agent.
3. It should be possible to blame this agent for his illegal act.

Retributivism If a mother commits theft, the theft is illegal and can be attributed to the mother (she did it). Under normal circumstances the mother can be blamed for what she did, and in that case the amount of blame is often seen as indicative for the amount of punishment. Little blame should lead to little punishment, much blame to much punishment. This view, according to which the amount of punishment should reflect the amount of blame, is called retributivism. If the mother was forced by a kidnapper who holds her children, she cannot be blamed for what she did. Therefore, under classical criminal law, the mother cannot be punished because she has no blame, although her act can still be attributed to her.

No punishing anymore What would change under hard incompatibilism? The first condition for punishability remains unaffected because the illegality of an act has nothing to do with free will or with the presence of a traditional self.[53] If hard incompatibilism does not go as far as refusing to

---

[52] Notice that here the idea of the vanishing self combines with the idea of determinism. The self has been analyzed in terms of brain processes, and these brains processes are arguably determined by causal laws.

[53] Perhaps hard incompatibilism would also exclude the possibility to speak about actions, rather than mere events, and in that case it would not make sense to speak of the illegality of an act either.

recognize the very existence of acts and agents – and there is no need for it to do so – the second condition remains unaffected too. However, the third condition cannot be satisfied anymore, since it is only possible to blame an agent for an act for which the agent can be held responsible.

We see that under hard incompatibilism the third condition for punishability cannot be met. Should we therefore draw the conclusion that under hard incompatibilism nobody can be punished anymore and that penal law should be abolished? Yes, that should be the conclusion if punishment is defined in a narrow way, as involving a sanction that was *deserved* by some crime. If there is no responsibility, nobody deserves to be sanctioned.

Another way to make the same point has to do with the phenomenon on non-accountability because of insanity. Insanity is usually considered to be a reason why somebody who committed a crime should not be punished, or should receive a diminished amount of punishment. On hard incompatibilism the distinction between sane and insane persons loses most of its relevance. Nobody is responsible, and *in a sense* every criminal should be treated as an insane person. Not because the fact that somebody committed a crime proves him to be insane, but because it has become senseless to attribute responsibility to anyone. Hard incompatibilism makes punishment as retribution a senseless affair.

Punishment as measure However, it is also possible to take 'punishment' in the broader sense of a measure that is taken against some person for the reason that he acted illegally. For example, a murderer may be incarcerated, not to make him suffer for what he did, but to prevent him (and perhaps also other potential murderers) from committing new crimes. Whether that should be done, and if so, under what conditions, is a matter that requires serious study, but taking preventive measures against a person does not conflict with hard incompatibilism. In this connection the difference between sane and insane persons might become relevant again, because sane persons might benefit from some sanctions which would be without effect for insane persons. (See Section 8.3)

Diminished capacity We have seen (in Section 2.3) that a court may recognize the phenomenon of diminished capacity, in case an agent's capacities were impaired by some mental disorder, but not to an extent of full legal irresponsibility. The recognition of such diminished capacity may lead to a mitigation of the punishment.

From the perspective of a hard incompatibilist, this practice does not make much sense. A punishment may be useful if its object is capable to learn from it, but then there is no reason to mitigate the punishment. The amount of punishment should be determined by what is necessary to obtain the desirable effects that are aimed for. If punishment is not useful because the person who is punished will not learn from it, there is no reason to impose a limited amount of punishment. Even this limited amount will do no good.

The phenomenon of limited punishment in case of diminished capacity can only be explained from the idea that somebody with diminished capacity can be blamed a little less and therefore deserves a little less punishment. For a hard incompatibilist, who does not believe in blame or in deserved punishment, this explanation only illustrates the wrong presuppositions underlying our actual social practice of punishing.

Allowed sanctions In general, all 'sanctions' that aim at bringing about some future effects, including prevention of crimes, giving satisfaction to the victims of the crime, and rehabilitation of the criminal, are compatible with hard incompatibilism. Excluded are all sanctions that are somehow connected to

the blame attached to the criminal because of his crime. Excluded are also limitations on sanctions because they would exceed the blame that can be put on the criminal. By definition for a hard incompatibilist there is never any ground for blame and therefore the amount of blameworthiness cannot function as a limit on the sanctions.

Another consequence that many will find shocking is that there seems not even to be reason anymore to limit the sanctions to persons who committed a crime. Those persons do not differ from other persons in being blameworthy or deserving punishment. Nobody deserves punishment and in that respect criminals and non-criminals should be treated in the same way. However we will see (in Section 8.3) that it may nevertheless make sense to distinguish between criminals and non-criminals.

8.2 Incompatibilism and liability in private law

Fault liability Private law deals with the mutual relations between legal subjects. One of the topics that private law deals with is liability for damages. If somebody suffers damage, for instance bodily harm, or loss of or damage to some of his goods, the main rule is that he must bear that damage himself. In principle nobody should be made to pay for somebody else's damage. However, there are exceptions to this principle; sometimes people are made to pay for the damage of somebody else, and the main reason why people are made to do this is that they caused the damage and that they can be blamed for having done so. A typical example would that a car driver culpably causes a collision with another car. The driver is liable for the damage both to the driver and passengers of the other car and to the other car itself. In such a case we speak of fault liability. Somebody is liable for the damage that was caused by his fault, where a fault is an act that was both unlawful and can be attributed to an agent who deserves blame for having committed the act.

If somebody causes damage to somebody else but cannot be blamed for it, the main principle is that is that the liability for the damage is not shifted to the person who caused it, since that person could not help it. For example Jan takes drugs which normally have no serious side-effects. However, unforeseeably the drugs cause an epileptic attack. During this attack Jan damages expensive goods in an antique shop he was visiting. In principle Jan is not liable for the damage of the shop owner, because he could not foresee the attack and did not have to take the possibility into account.

Strict liability The problem with cases such as the one about Jan who had an epileptic attack is that the damage remains with the shop owner, who could not help the damage either. Therefore private law sometimes shifts the damage from the person who suffered it in the first place to the person who caused it, even though the latter person could not help it. The argument is then more or less as follows. Some damage occurred and nobody is to be blamed for it. The law must determine who has to bear the damage. Normally this will the person who suffered it in the first place, but there may be situations in which the damage should fall within the risks of somebody else. For example, the owners of defective goods that caused damage are often liable for the damage, even though they could not help the damage. In such cases, where somebody can be liable even though he could not be blamed for the damage, we speak of strict liability.

Suppose that hard incompatibilism is the correct view on responsibility. Then it is not possible to blame anybody for damage that was caused. The idea of fault liability does not make sense then, because nobody will ever have committed a fault. This means that the considerations that introduced strict liability in private law become relevant in a vastly wider domain of cases. Damage occurs and it is a matter of policy who has to bear the damage.

Reasons for shifting liability Ideas that presently play a role in shifting damage from the person who suffered it in the first place to somebody else belong to at least four categories. One reason to shift the damage has to do with fairness. For example, the introduction of cars into everyday life has created the risk of people being injured or even killed in a car accident. Although such accidents are sometimes to be blamed on the victims themselves, it cannot be denied that the use of cars enormously increases the chance of serious traffic accidents. The imposition of strict liability on car owners, who profit from the use of cars, means that the resulting damage does not have to be borne (completely) by the victims of these accidents.

Another reason to impose strict liability has to do with economic efficiency. One way to distribute damage over society may lead to less overall costs than another way of distributing the damage. For instance, it is easier and probably cheaper that the owners of tigers insure themselves against the risk that the tigers will escape and wound people, than it is for ordinary citizens to insure themselves against the risk that they will be wounded by a runaway tiger. Imposing a strict liability on owners of dangerous animals such as tigers will then lead to less costs for society as a whole.

A third function of strict liability is to allow victims who suffered damage caused by the behavior of someone else or because of an event for which the victim was not responsible himself to recover this damage. This function explains both strict liabilities and liabilities for damage causing behavior of other persons, including the liability of employers and parents for their employees respectively children. For example, if employers are liable for their employees and parents for their children, this increases the chance that the victims of damage caused by employees and children will receive compensation. This chance is even further increased if employers and parents are obligated to insure themselves for these liabilities.

A fourth function of strict liability is to prevent the occurrence of damage causing events. By making persons other than those who actually suffered the damage liable to compensate for this damage, tort law promotes that these other people will be more careful to avoid damage. For example, a car producer who must compensate for the damage caused by a defective car will be stimulated to invest even more in security checks.

These four functions of strict liability become even more important if fault liability is abandoned in favor of strict liability. As we will see in the next subsection, the prevention of damage may even ask for the re-introduction of some kind of fault liability.

8.3 Soft incompatibilism

As the reader may have noticed, it is very difficult to abandon the traditional view that there are persons who perform actions, who are normally responsible for their actions, who can often be blamed for what they did and who should suffer the consequences of their actions, in particular if they could be blamed for them. Even if we try to imagine what the competing brain processes theory of mind would imply, we tend to think about every individual process in the mind as if it were a complete individual person. That does not prove that the traditional view is correct, but it is a reason to take into account that most people will continue to ascribe actions to agents and to hold agents responsible for what they did. Under hard compatibilism it does at first sight not make much sense to attribute actions to agents and to hold these agents responsible. What would make sense is to attempt to influence future events, including future actions. We see this in criminal law, if 'punishment' is used to prevent future crimes, to satisfy the victims of a crime, or to rehabilitate the

criminal. We also see it in private law when strict liability is used to prevent future damage or to bring about a more fair distribution of damage over society.

Hard incompatibilism leads to legal policies which are directed at the future and do not react to what happened in the past, in particular not where the attribution of responsibility is concerned. However, in bringing about the best possible future results it may be opportune to take into account that many people will continue to live in a world where people hold each other responsible for their acts. Perhaps this world view, wrong as it may be, should be incorporated into the legal policies concerning crimes and damage causing events (actions). If a person knows that he will be punished for what he does, that can influence his future behavior. This still holds if the punishment is wrongly characterized as something which the criminal deserved for what he did. Backward looking this may not make much sense, but forward looking it may be effective in diminishing the amount of crime.

Punishing as if … It was argued above (Section 8.1) that at first sight incompatibilism would not distinguish between sane and insane persons, or between criminals and non-criminals for the issue whether to 'punish'. Neither sane nor insane persons, neither criminals nor non-criminals would be responsible for their actions. The only thing that would count would be the future effects of the measures that are taken under the moniker of 'punishment'. However, this conclusion would be too hasty in the light of the fact that most people will continue living as if sane people can sensibly be held responsible for what they do, while insane people cannot, and as if criminals deserve to be punished while innocent people do not. The effects of punishing insane persons would be different from the effects of punishing sane persons. Perhaps a sane person can be improved through 'punishment', while an insane person would not profit. 'Punishing' innocent persons might create upheaval in society, while 'punishing' criminals would not. Similar points can be made regarding the question whether it is good to shift liability from the person who suffered the damage in the first place to somebody else. People might find it unpalatable if damage were too often shifted to 'innocent' persons.

Therefore it might be profitable to have rules for criminal and private law liability that somehow take the 'responsibility' of persons into account. This might even be the case if hard incompatibilism were true. It might be the case that the best law for a society in which nobody can be held responsible would distinguish between people who are 'responsible' and people who are not. The resulting approach might be called soft incompatibilism.[54]

9.   A schematic overview

We have almost reached the end of our argument, and it may be useful to recapitulate some of the major issues. The mechanistic view of mind, exemplified in both causal determinism and the competing brain processes theory challenges the traditional image of man that underlies our actual social practice of attributing responsibility and assigning liability. There are two main ways to deal with this challenge: (1) compatibilism and (2) incompatibilism (in a broad sense).

Compatibilism holds that the mechanistic view of mind is by and large compatible with our present social practice of holding people responsible for their doings. There at least two ways to defend this position, be it that they both depend on the same underlying strategy of holding that some characteristics of agents are the result of attribution (see Section 6.4):

---

[54]   This 'soft incompatibilism' should not be confused with 'soft determinism', an expression that is sometimes used in the literature to denote compatibilism.

Strategy 1: admits that possibly agents lack free will, but assumes that in our social practice we do not need free will to attribute responsibility.

Strategy 2: requires free will for the existence of responsibility, but assumes that free will does not depend on causal determinism but on the presence of certain *capacities*; these capacities are the result of attribution.

Incompatibilism in the broad sense is the denial of compatibilism: the mechanistic view of mind makes our present social practice of holding agents responsible for their doings *in principle* obsolete. This form of incompatibilism does not automatically imply that our present social practice is obsolete.

Libertarianism One reason why that is not the case would be that the mechanistic view of mind does not block free will. There are again two variants of the libertarian view:
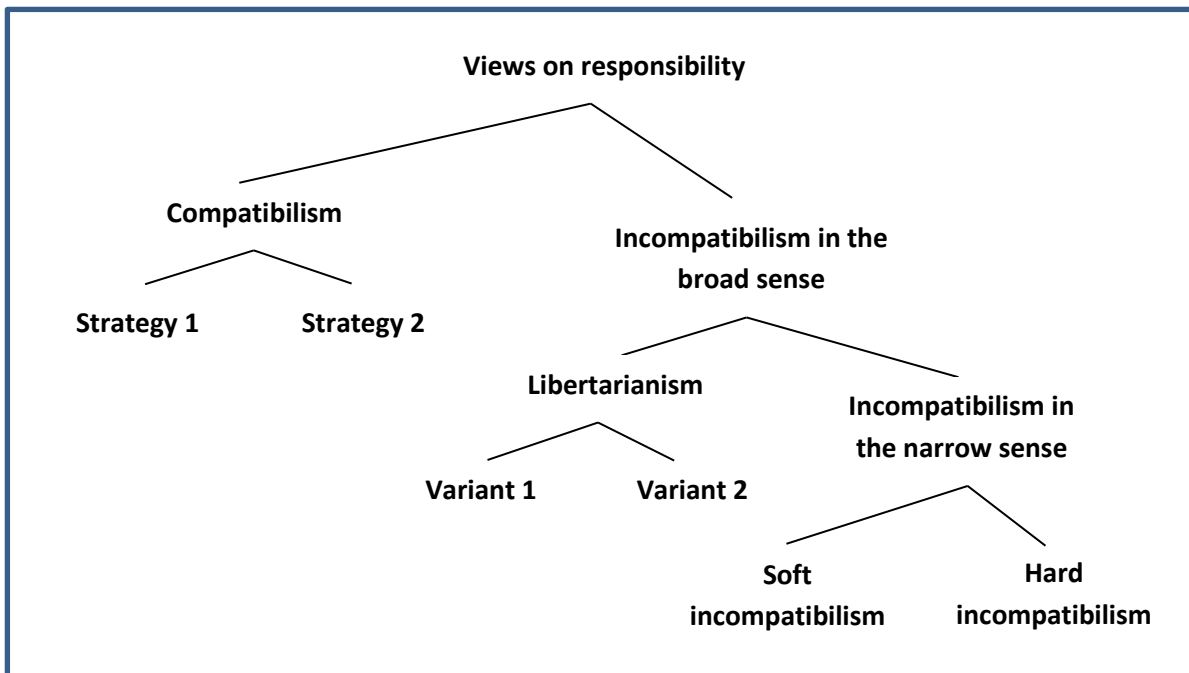
Variant 1: Causal determinism does not hold in general, and therefore it does not hold for human decision making either. And because causal determinism does not hold for human decision making, the human will is free. And because the human will is free, it makes sense to hold agents responsible for their doings.

Variant 2: In general causal determinism holds, but the human mind is an exception, because the determinism on brain level does not translate into determinism on the level of mind. Therefore the human will is free and therefore our practice of holding agents responsible makes sense.

Incompatibilism in the narrow sense is the view that there is no free will and that therefore our social practice of holding agents responsible for their doings *in principle* does not make sense.

Soft incompatibilism and forward-looking responsibility In principle, because- as we have seen above – it may still make sense to hold people responsible for their doings, for instance because this leads to the best results for the future. Responsibility is then primarily a future-oriented notion: we hold people responsible because of the attractive consequences this has for the future. This is the view that was labeled 'soft incompatibilism' above. Notice that soft incompatibilism, being a variant on incompatibilism in the narrow sense, considers our backward-looking practice of holding agents responsible senseless. Agents do not deserve praise or blame for what they did, although it may be useful to praise or blame them with an eye to the future.

Hard incompatibilism is also a variant on incompatibilism in the narrow sense. It considers responsibility not so much as something that we attribute to agents, but rather as a characteristic that agents have because of what they did. Given this meaning of the word 'responsibility', the only factor that determines whether an agent is (not: is held) responsible is whether he freely did what he did. Since free will does not exist on this view, no agent ever acted freely, and therefore no agent ever was, or will be, responsible for what he did.

```
                        Views on responsibility



      Compatibilism
                                        Incompatibilism in the
                                             broad sense

  Strategy 1       Strategy 2
                                    Libertarianism
                                                        Incompatibilism in
                                                        the narrow sense

                              Variant 1        Variant 2

                                                    Soft              Hard
                                               incompatibilism   incompatibilism
```

10.   Two perspectives

We have seen that the practice of holding people responsible for their doings and the legal extension of this practice of punishing persons who can be blamed for some crime and of shifting liability to persons to whom the damage can be blamed, are based on a particular image of man. Basically this image boils down to a rational self that is the collection point for all the information that somebody receives through his senses, and the point of departure of all acts performed by this person. This self is the free source from which all acts spring.

This traditional image is not well compatible with the image of man offered to us by modern physical science. According to this 'scientific image', there may not be a single self, a person does not (always) act on his conscious decisions, and it is not even understandable what a free will might amount to, let alone that we would have one.

Is one of these two images correct, and should we adapt our legal provisions to the correct image? Or should we allow the two images to co-exist independently and continue to practice law on the same footing as it used to have? There is something to be said for the latter approach. The 'scientific image' of man is only a limited image that focuses on man's physical characteristics, possibly to the detriment of man's non-physical characteristics. Physical science seems not to be the best possible tool to deal with the aspirations of mankind, man's hopes and dreams, his fears and anxieties, his love and repulsion. A more humanitarian approach might do better justice to all these phenomena which to a large extent govern the life of man. Perhaps the law should continue to be based on such a humanitarian outlook.

However, the image of man sketched by the physical sciences seems to be also 'correct'. Our theories framed in the terminology of physics and chemistry is more and more successful in predicting the phenomena that pertain to our bodies. Should we ignore this successful research program to stick with a time-tested but also time-worn humanitarian tradition? Suppose that both

the humanitarian and the physical perspective on man are both correct. Then we should expect that these two perspectives lead to compatible results. If we look at a house from two different sides, the images will be very different. However, we assume rightly that the two images are both only partial and that it should be possible to join them in one more complete image. The adoption of the one image should not make us reject the other image.

In this connection a parallel may be useful. According to a traditional view of the world, the sun rises every morning and sets every evening. The old Greeks even had a story to explain this. The god of the sun, Helios, drives his chariot containing the sun in an orbit around the Earth, and in the morning we can see the wagon climbing the sky, while in the evening we can see it descend again. The modern scientific view holds that it is not the sun that rises, but the earth that revolves around its axis that explains our visual phenomena. The scientific theory at once refutes the traditional view – it is not the motion of the sun that explains what we see – and confirms it, because what we see can be explained from the theory that earth rotates around its axis. The two views are in a sense alternatives for each other, but is some deeper sense complement each other. The modern scientific theory explains why people originally held the traditional theory, while the traditional theory is more in line with the way we experience the world around us.

Perhaps we should take the same approach to the humanitarian and the physical perspective on mankind. Both perspectives are in development and can be improved in the course of time. Ideally they should in the end become compatible, both recognizable as an incomplete but nevertheless correct image that can be combined with the image of the other discipline. We should not abandon the humanitarian image of man and replace it with the image of the physical sciences. However, we should neither anxiously cling to the traditional humanitarian image, and base our law on it while ignoring the insights from modern science. We should attempt to combine both images of mankind into a coherent more encompassing image, and be willing to adapt both images in order to make the combined coherent image possible. [55]

References

Boaz, David (ed.) (1998), *The Libertarian Reader* (New York: The Free Press 1998).

Brian, I. and A. Chatterjee (2004), 'Alien Hand Syndrome', in Archives of Neurology 61 (2004), 292-294

Churchland, Paul (1994), 'folk psychology (2)', lemma in Samuel Guttenplan, *A Companion to the Philosophy of Mind* (Oxford: Blackwell 1994), 308-316.

Davidson, Donald (1963), 'Actions, Reasons and Causes', in *Journal of Philosophy* 60 (1963), 685-700.

Dworkin, Ronald (2011), *Justice for hedgehogs* (Cambridge (Mass.): The Belknap Press of Harvard University Press 2011).

Eckhardt, Barbara von (1994), 'folk psychology (1)', lemma in Samuel Guttenplan, A Companion to the Philosophy of Mind (Oxford: Blackwell 1994), 300-308.

Hage, Jaap (2009). 'The Meaning of Legal Status Words', in Jaap Hage and Dietmar von der Pfordten (eds.), *Concepts in Law* (Dordrecht ea.: Springer 2009), 55-66.

---

[55] An extended argument why such an integration is possible and desirable is Anne Ruth Mackor, *Meaningful and Rule-Guided Behaviour: A Naturalistic Approach*, PhD-thesis Rijksuniversiteit Groningen 1997.

Healey, Richard (2009), 'Causation in Quantum Mechanics', in Helen Beebee ea. (eds.), *The Oxford Handbook of Causation* (Oxford: OUP 2009), 673-686.

Keiler, Johannes and David Roef (eds.) (2015). *Comparative Concepts of Criminal Law* (Cambridge ea.: Intersentia 2015).

Kripke, S.A. (1972), 'Naming and necessity', in D. Davidson and G. Harman (eds.), *Semantics of natural language*, 2nd ed., Reidel, Dordrecht, 253-355.

Kuhn, Thomas S. (1962). *The Structure of Scientific Revolutions*, Second Edition, Chicago: The University of Chicago Press 1970, (1st ed. 1962).

Lakatos, I. (1970). 'Falsification and the Methodology of Scientific Research Programs', in I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, (Cambridge: Cambridge University Press), 91-196.

Lamme, Victor (2013), *De vrije wil bestaat niet* (Amsterdam: Prometheus. Bert Bakker 2013).

Libet, C.A, Gleason, E.W. Wright and D.K. Pearl (1983), 'Time of conscious attention to react in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act', in *Brain* 106 (1983), 623-642.

Mackor, Anne Ruth (1997), *Meaningful and Rule-Guided Behaviour: A Naturalistic Approach*, PhD-thesis Rijksuniversiteit Groningen 1997.

Mackor, Anne Ruth (2013). 'What can Neurosciences Say about Responsibility? Taking the Distinction between Theoretical and Practical Reason Seriously' in Nicole A. Vincent (ed), *Legal Responsibility and Neuroscience, Series on Neurosciences and Law* (Oxford: Oxford University Press, 2013), 53-83.

Martin, Michael and Lee C. McIntyre eds. (1994), *Readings in the Philosophy of Social Science* (Cambridge Mass.: MIT Press 1994).

McLaughlin, Brian (1994), 'epiphenomenalism', lemma in Samuel Guttenplan (ed.), *A Companion to the Philosophy of Mind* (Oxford: Blackwell 1994), 277-288.

McLaughlin, Brian , Ansgar Beckermann and Sven Walter (2009), *The Oxford Handbook of Philosophy of Mind* (Oxford: OUP 2009).

Moore, G.E. (1903). *Principia Ethica*. (Cambridge: Cambridge University Press 1903).

Morse, Stephen J. 'Rationality and responsibility', *Southern California Law Review*, 2000, 251-268.

Pereboom, Derk (2001), *Living without free will* (Cambridge: Cambridge University Press 2001).

Raz, Joseph (1975), *Practical Reason and Norms* (London: Hutchinson 1975).

Ridge, Michael (2014), 'Moral Non-Naturalism', *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2014/entries/moral-non-naturalism/>.

Rosenthal, David M. (1994), 'identity theories', lemma in Samuel Guttenplan (ed.), A Companion to the Philosophy of Mind (Oxford: Blackwell 1994), 348-355.

Sie, Maureen and Arno Wouters, 'The BCN Challenge to Compatibilist Free Will and Personal Responsibility', *Neuroethics* 3 (2010), 121-133.

Smits, Jan (2015), 'Does the law of delict have a future? On neuroscience and liability', *Maastricht European Private Law Institute Working Paper*, No. 2015/1, p. 1-10.

Verplaetse, Jan (2011), *Zonder vrije wil* (Amsterdam: Uitgeverij Nieuwezijds 2011).

Vincent, Nicole A. (2009), 'On the Relevance of Neuroscience to Criminal Responsibility', in *Criminal Law and Philosophy* 2009.

Walter, Sven (2009), 'Epiphenomenalism', in McLaughlin, Beckerman and Walter, *The Oxford Handbook of Philosophy of Mind* (Oxford: OUP 2009), 66-84.