Torgeir Støen

You can do what you want, but can you want what you want?

Free will as a concept in forensic psychiatry

Graduate thesis in Clinical programme of psychology
Supervisor: Robert Biegler
Trondheim, November 2016

Norwegian University of Science and Technology
Faculty of Social Sciences and Technology Management
Department of Psychology

NTNU
Norwegian University of
Science and Technology

The general delusion about free will obvious …

One must view a wrecked man like a sickly one ... It is not more strange that there should be necessary wickedness than disease.

Charles Darwin

Foreword

First, I would like to thank my supervisor, Robert Biegler, for his guidance and patience throughout this project. Robert, you have been an inspiration to me. In addition, I would like to thank Mari and Rannveig for proofreading this text, and for their support, as well as André, Ida, Ann-Karin, and finally my mother, Kirsten, for all support.

Abstract

This thesis has investigated the concept of free will, and its role in forensic psychiatry. More specifically, this text discusses common definitions of free will, to see whether any of the definitions are coherent and testable. It is argued that, although there may exist sound definitions of free will, the most commonly used definitions are incoherent. Historically, free will has been tied to the notion of moral responsibility and retributive philosophies of punishment. Legal scholars have noted that free will remains an important part of justice systems, insofar these systems are based on retributive reasons for punishment. Traditionally the academic discussion regarding responsibility has focused on how and when responsibility should be ascribed. However, in this thesis it is argued that the legal understanding of responsibility itself might not be sound, given that it seemingly depends on an unsound conceptualization of free will. This seems to challenge the legitimacy of the current state of forensic psychiatry, and the current way of ascribing criminal responsibility, and punishment.

Innhold

1.0 Introduction

In the wake of the trial against Anders Behring Breivik, an extensive principal discussion regarding the notion of legal insanity was raised. A committee was appointed to investigate whether the judicial system should have rules concerning criminal accountability, and how these may be formulated (NOU 2014:10). Key issues raised were which criteria were to be applied for clearing a perpetrator of moral responsibility, and on the other hand, why others should indeed be regarded as morally responsible for their actions.

The notion of moral responsibility proves to be a complex issue, but many authors argue that it relies on the notion of "free will" and whether we have it (Cashmore, 2010; Greene & Cohen, 2004; Moore, 2012; Syse, 2014). The free will-debate has been ongoing for millennia (Dennett, 1984; Waller, 2011, 2014), and most, if not all, prominent philosophers have had something to say about it. Recently, the debate has regained interest, not just in the field of philosophy, but also within the scientific community. For instance, Benjamin Libet took an experimental approach to the question of free will, showing that voluntary acts are preceded by specific electrical changes in the brain that begins about 550ms before the act (Libet, Gleason, Wright, & Pearl, 1983). He later concluded that volition is unconsciously initiated, but that "free will" may still be able to veto acts, and therefore was not disproven (Libet, 1999). Cashmore (2010) argues that biologists may have paid too little attention to the apparent lack of free will, urging the scientific community to pay attention to the question, proposing that it may have serious implications for the concept of moral responsibility.

Many authors and philosophers would claim that the existence of free will and moral responsibility is related to substance dualism (Monroe, Dillon, & Malle, 2014), that it is dependent on whether we live in a deterministic or indeterministic universe (Ebert & Wegner, 2011; Latorre, 2013), while others argue that free will is compatible with physical reality regardless of whether determinism is true (Dennett, 1984), again, others say that it is *incompatible* regardless (Cashmore, 2010; Pereboom, 2001; Pierre, 2014). A rare position is that the system of moral responsibility should be abolished, while free will should be redefined as a psychological construct without the common link to moral responsibility (Waller, 2011, 2014). Lastly, some may claim that the debate is exhausted, arguing for a pragmatic approach to the problem: Whether we have free will, and are ultimately responsible

for our actions or not, is irrelevant; the belief in the current system of moral responsibility is beneficial for our society, and we should therefore keep this worldview (Juth & Lorentzon, 2010; Smilansky, 2000).

In Norwegian law, much of the reasoning behind punishment seems to be somewhat similar to the latter viewpoint. It is stated that retribution "cannot be the reason for punishment", but that it is still central to the doctrine, "mainly as a limiting factor" (NOU 2014:10 2014, p. 46), in other words; that all are regarded as morally responsible *by default*. Aslak Syse concludes that:

> The Breivik case nevertheless demonstrates that retribution still remains as a justification for a penal verdict. An underlying element in many of the reports is that Breivik should be found fit to stand trial, so that society can give him «the punishment he deserves (Syse, 2012, p. 841).

Thus, to some extent it is a *retributive*, as opposed to a pure utilitarian, consequentialist doctrine of punishment. In the report, it is stated that the premise of free will is not explicitly considered as a condition of liability, as it relates to a metaphysical entity that may hardly be explained scientifically (NOU 2014:10). Worth mentioning is that the committee does not rule out that the penal theory itself may rely on the philosophical assumption of indeterminism. Although labeling such challenges as *radical*, it is concluded in the 440 page report that it *"may be difficult to maintain an action- and guilt-oriented accountability doctrine if it were to be proven that individuals lack free will"* (NOU 2014:10, pp. 424-425, my translation). Hence the notion of free will remains an integrated part of Norwegian criminal law (Christie, 2007; Syse, 2006, 2014); penal sanctions may depend on it *"insofar they are meant to harm the perpetrator"* (Syse, 2014, p. 398).

There are objections to many of the claims made by the committee. First, given that few of our scientific models regard any part of reality as 'free' from causality, one could wonder on whom lies the burden of proof. Secondly, the view that the free will debate is highly dependent on the philosophy of metaphysics, as opposed to being just as much a subject to scientific rigor as any psychological construct, may also be called in to question. Regardless

of what implications, if any, which may follow scientific truths about free will, there still seems to be a need for a stronger consensus on the subject.

Due to the apparent lack of such a consensus, the Norwegian judicial system has decided to operate according with the current penal law. This involves an assumption that, even though it may not be consistent with reality, it is in any case beneficial to assume that most of us at least to some degree are *free* to choose our actions and therefore morally responsible, while a few others are not (NOU 2014:10). The accountability doctrine of the Norwegian judicial system can stand in contrast to the Swedish system, which rejects it (Juth & Lorentzon, 2010; NOU 2014:10, p. 65). Comparing the Norwegian judicial system (or even most other judicial systems) with the Swedish system may shed light on the fact that ideas about free will could affect views on criminal responsibility, and the principles that guide sanctions and punishment. Nevertheless, Juth and Lorentzon (2010) argue that even though views on free will historically have had an impact on judicial systems and the perception of moral responsibility, they should not.

In line with recent contributions (Cashmore, 2010; Libet, 1999; Pierre, 2014), I will try to shed light on why the free will-debate should not be contained within the hermeneutics of a pure philosophical approach, and that it may very well be subject to empirical research fields such as psychology, biology and neuroscience.

In this text, I will question whether there is anything we may call "free will" within a scientific framework. In doing this, it is important to avoid begging the question, as there are probably nearly as many definitions of the term as there are authors discussing it. Trying to find coherent definitions is therefore paramount, if one wants to contribute to the debate. I will aim to find the most common definitions of free will, and see which are clear enough to be tested against empirical data.

In conclusion, I will try to answer the following questions:

1) Are there any coherent and testable definitions of free will?
2) How does the various definitions relate to the conception of moral responsibility and retributivist legal theory?

3) What might be the possible implications of rejecting the free will construct in a legal and scientific context?

1.1 The confusion over deciding Breivik's legal sanity

Two consecutive terror attacks on July 22, 2011 left an entire nation in shock, as Norwegian Anders Behring Breivik killed 77 people, many of whom were children and adolescents. At 15:25 8 persons were killed, and 9 were injured by the detonation of a 950kg fertilizer car bomb in the government quarter. Two hours later, Breivik had managed to travel directly from the bombsite to the small island Utøya, where the summer camp for the Norwegian Labour Party youth organization was situated. Masquerading as a police officer responding as a security measure to the recent bombing, he was granted access to the island ferry. Upon arriving the island, he almost immediately started shooting towards the 600 participants of the youth camp. Bone-chilling witness reports describe Breivik laughing and shouting while performing executions with hollow point bullets, and persuading children to come forward while identifying as a police officer. A further 69 persons were killed, 59 of whom were born after 1990. After 50 minutes, he called the police to surrender, identifying himself as "Commander Anders Behring Breivik from the anti-communist resistance movement". Hours prior to the attack, he had distributed a 1,518-page so-called Maniphesto describing reasons behind, and preparations prior to the attacks (Berntzen & Sandberg, 2014; Melle, 2013).

After the subject was apprehended, there were good reasons to question his sanity. The court appointed two expert medical witnesses to conduct a psychiatric evaluation. The psychiatrists characterized Breivik's beliefs and actions as "bizarre" and ideas as "part of a bizarre delusional system" (Moore, 2014). According to DSM-IV, *bizarre* delusions are beliefs that are clearly implausible, not understandable and which do not derive from ordinary life experiences (American Psychiatric Association, 2000). They concluded that Breivik was 'psychotic' while planning and committing his acts, as well as during the psychiatric evaluation, therefore not legally accountable for his actions. This process attracted intense media coverage, and the conclusion lead to prominent newspapers and politicians demanding a new psychiatric evaluation (Melle, 2013). The court appointed a second pair of psychiatrists who came to the opposite conclusion about his legal sanity. They found that Breivik had "no bizarre delusions" of phenomena that are physically impossible or intelligible, even if they were extremely unrealistic and wrong. They concluded that he was not psychotic during the

examination nor when he was performing his crimes, and therefore not legally insane (Moore, 2014).

Clearly, the question of whether Breivik was to be held morally responsible for his actions was a complex issue, but the reasons for this decision seemed somewhat arbitrary, seemingly to a large degree based on the intelligibility and plausibility of his beliefs. Nevertheless, it seemed to be the deciding factor in a dichotomous choice, with significant implications: either the defendant is sane and therefore morally responsible, and should be *punished* by prison, (or in some jurisdictions, even death). Alternatively, he is legally insane, and therefore not morally responsible, and should be *excused*, and, in his case forcibly *treated* for his illness at a medical hospital. Since the latter alternative would mean that a perpetrator is *not* morally responsible, the level of hazard must be high to justify forced hospitalization, which requires 'adequate resources', 'positive and meaningful facilities', and 'exert no limitations to freedom other than necessary for the protection of others' (NOU 2014:10, p. 33).

Oslo University Hospital reported that if Breivik were convicted to treatment, they would have had to hire 20 people to work in shifts exclusively with him: two shifts of four people in the daytime, and three people during the night shift. Additionally, a psychiatrist would have to be on call. Note that it is illegal for hospitals to lock in their patients during the night, which is the reason for the high staffing 24/7. The University Hospital calculated the total costs of treating Breivik to be somewhere around 20 million Norwegian kroner per year. In contrast, Ila prison facility reported that the additional security costs for keeping Breivik would be around 4,6 million (Steen, 2012). Of course, locking in *prisoners* during the night is legal.

So then how can we justify such divergence based on what, in Breivik's case, seems to be an arbitrary common sense evaluation of the content of his beliefs? This is where the 'limiting factor' of retribution might come in to play: Moral responsibility is the *default* position in our judicial system: Most people are morally responsible, and if they decide to commit a wrongdoing, they *deserve* whatever current conditions of a prison sentence, unless 'proven' that they are not, and must be *excused* from the wrongdoing as a 'special measure' (Syse, 2014). In fact, a special high-security mental health unit was built within a prison facility *especially for Breivik* while the trial was ongoing. Later on, this unit never opened, because he

was *not* found legally insane, and therefore could be sentenced to a high-security prison unit instead (Syse, 2014). Clearly, there are significant economical and humanitarian implications tied to this question. So why, and on what grounds do we make the distinction between the deserving and the excused?

1.2 Norwegian penal code

The Norwegian penal code states four cumulative conditions for moral or judicial responsibility (NOU 2014:10; Syse, 2006, my translation):

a) The criminal act is punishable by law
b) There can be no sufficient causes for impunity (e.g. self-defense)
c) The condition of mens rea – The perpetrator has to have acted with a sufficient degree of intent, principally in the moment of action
d) The condition of legal sanity.

The question of mens rea could be understood as whether the consequences of the wrongdoing were intentional. For instance, if someone suffered a concussion, and later died from a brain aneurism after being punched in a pointless drunken fistfight, should this subsequent fate of the victim affect the offender even when the offender did not even remotely intend or expect the outcome? In other words: Should the offender be convicted of murder for something that otherwise potentially could have been solved by an apology?

1.3.1 Moral responsibility

Responsibility is often accompanied with a qualifier, to create terms like “personal responsibility”, “criminal responsibility” and “moral responsibility”. Often, these terms may have been used interchangeably. Although I intend to show how and why it may be important to make distinctions as we use some of these concepts, the terms have all tended to be viewed as axiomatically tied to-, or inseparable from some conception of free will (Fischer, Kane, Pereboom, & Vargas, 2009; Waller, 2014).

1.3.2 Moral responsibility and Legal insanity

In legal theory, traditionally as well as currently, responsibility is conceived similar to that in Kantian cognitivist ethics (Thorvik, 2000, 2008). According to Kant’s philosophy, the human

being has a natural *faculty* of *understanding*, and is able to *think of* himself as an "I" instead of just *feeling* himself. This makes him a *person* and thus '*infinitely* above all other living things on earth' (Kant, 2006, p. 15, my emphasis). Accordingly, this unique reasoning capacity makes the 'person' an entity which is an entirely different being from *things* such as irrational animals and inanimate objects and hence ultimately responsible for its actions. In short, this is how the Kantian, '*ought implies can*'-link is enabled: If a human agent 'knows' he is morally obliged (ought) to do something, he must also be capable (can) under 'natural conditions'. In these instances, breaking such obligations may and should be punished, according to Kant, even by death in some cases.

Insanity is however addressed in Kant's doctrine: When influenced by a certain set of conditions, like young age, mental illness, dementia, or mental retardation, the human agent may be 'blocked´, and may lack the presumed unique, but nonetheless 'natural' given ability to reason 'properly´ and to act according to right and wrong. In these instances, a perpetrator must to be met with understanding because of the apparent restraints on his mental capacities, and thus can be excused from responsibility (Thorvik, 2000).

Although it is more than two centuries old, the Kantian doctrine of moral law nonetheless appears remarkably similar to modern legal theory. In Norwegian legal theory, the notion of personal blame and responsibility for a crime rests on the notion that the perpetrator is normally assumed to be free in his agency, meaning he should have *and* could have done otherwise than he did in regards to that crime (NOU 2014:10). Even if the first three conditions for judicial responsibility are met (section 1.3), the perpetrator may not be punished for a crime if he is deemed legally insane at the moment of action, thus not meeting the fourth condition (Syse, 2006). Legal insanity depends the following sort of conditions: Firstly, the perpetrator is not legally sane, and not accountable, if he is less than 15 years old, "psychotic", mentally retarded, or is under a "strong disturbance of consciousness" (Syse, 2006).

All civilized societies have rules considering the sanity of a criminal perpetrator (NOU 2014:10). Perhaps there are certain individuals so mentally disturbed that they will not be deterred by- or benefit in any way from punishment alone, or so that some of us would want

to hold them accountable or 'blameworthy'. Still, there are indeed great challenges when one considers the rules for legal insanity. The main one may be to achieve accuracy, so that only those that 'should' are deemed legally insane (NOU 2014:10). In other words: to avoid false positives and negatives, whatever they could be. When forming a discussion regarding the grounds for these rules, it appears customary to consider two alternative overarching principles on which to ground the required assessment, namely the *psychological principle* and the *medical principle (Moore, 2014; NOU 2014:10; Rosenqvist, 2012).*

1.4 The Psychological principle

According to the psychological principle, the deciding factors for moral responsibility are the perpetrators mental abilities such as insight and ability to "choose freely" (NOU 2014:10, p. 47) as described in the section above. However, insanity tests based on the psychological principle regards *mental* illness itself to have a *weak relevance* to the notion of legal sanity, since it is defined as mental disease *plus some other factor* (Moore, 2014): In addition to the perpetrator having a mental disease or specific mental state, this condition may also be examined as a *causal factor* leading to the crime, for it to excuse the crime in question. For instance: If one knew that the reason that a paranoid schizophrenic killed a 7 year old, was a conviction the perpetrator thought that the victim was an immediate life threat, from a demon from hell or an alien from outer space, one could imagine that this would trigger empathy for the perpetrator. For if the defendant would otherwise never dream of intentionally harming a child, how could he possibly 'deserve', be deterred by- or learn from punishment alone, given that the victim of the crime was vividly perceived as something entirely different?

The M'Naghten rule has been paramount in Anglo-Saxon criminal law, and can be regarded as an expression of the psychological principle (Syse, 2006). The rule was developed as a reaction to the 1843 acquittal of Daniel M'Naghten, an Irish Protestant charged for the murder of Prime Minister Robert Peel's secretary, whom M'Naghten had mistaken for Peel himself. It appeared that the defendant was acting according to the conviction that he was persecuted by Robert Peel's political party, and by the Catholic Church. In light of these beliefs, the act could be regarded as justified, as the defendant feared for his life and thought he was acting under self-defense. The resulting M'Naghten rule states "all men are presumed to be sane and to possess a sufficient degree of reason to be responsible for his crimes, until the contrary be

proved to their satisfaction (NOU 2014:10, p. 73)". The rule has two branches, either of which is sufficient for exonerating the criminal from responsibility (Becker, 2003; Syse, 2006):

1. "*at the time of committing the act, the party accused was laboring under such a defect of reason, from the disease of mind, as not to know the nature and quality of the act.."*
2. *"or he did not know it, that he did not know he was doing what was wrong"*

Accordingly, M'Naghten was acquitted, viewed as constrained by a cognitive impairment and lack of understanding of the nature of- or unlawfulness of the act, and therefore was not exerting his free will (Syse, 2006; Thorvik, 2000). The first branch regards the defendants ability to have mens rea – which consists of mental states such as intent and premeditation (Moore, 2014). The second branch exonerates those who are incapable of considering the moral basis of the law and therefore cannot make moral choices (Becker, 2003). The M'Nagthen rule was criticized for its focus on cognition alone, which could be incompatible with our understanding of common mental disorders (Thorvik, 2008). For instance, both schizophrenia and mood disorders display both emotional and cognitive symptoms (World Health Organization, 1999).

To account for this problem, the **irresistible impulse test** could contribute by allowing for the possibility that the defendant could be excused if the criminal act was the *result of an irresistible impulse*, even if his cognition were not otherwise severely impaired (Becker, 2003, p. 43; Syse, 2006). Elements of the irresistible impulse test include

1. The perpetrator must have a significant mental illness.
2. The impulse must arise directly from the mental illness.
3. There must be no evidence of planning or premeditation by the defendant.

According to the irresistible impulse test, the lack of 'volitional capacity' as a result of disease, disorder or defect of the mind relieves the defendant from liability just as an impairment of 'cognitive capacity' (Becker, 2003).

Replacing the M'Naghten rule and the irresistible impulse test, the **Durham Test** includes a criterion of causality, and lists two *cumulative* conditions:

1. Did the defendant have a mental disease or defect?

2. If so, was it the reason for the act?

Becker (2003) points out two problems with the Durham Rule: First, the question of criminal responsibility is highly dependent on the expert medical witness, such as a clinical psychologist or a psychiatrist, in lack of a clear conception of what kinds of cases the law seeks to exempt from responsibility. Indeed the criteria for these diagnoses were postulated for the sake of treatment, not for the courtroom. Secondly, the question of causality is difficult, as it is difficult to separate the question of whether the act would have been committed had he not been mentally ill, from the question whether he was in fact ill when the act was committed.

The Moral Penal code, also called the *substantial capacity test* (Becker, 2003), was designed to avoid the problems of causation. Furthermore, the notion of absolute lack of knowledge of right and wrong in M'Naghten rule was changed to a more flexible *substantial capacity to appreciate* [right or wrong] (Syse, 2006). In addition, a volitional irresistible impulse-component was included. Thus only those lacking a "*substantial capacity to conduct or conform his conduct to the requirements of the law"*, as a result of *"mental disease or defect"* would not be held responsible. Nevertheless, given the flexibility and vagueness of this rule, those jurisdictions that used it, tended to fall back to a stricter practice similar to the M'Naghten rule (Syse, 2006). Cashmore (2010) points out that there is a significant degree of arbitrariness associated with this discussion.

Norwegian penal law rejects the causal criterion. However, a recent Norwegian study investigated rulings where the perpetrator has been found legally insane, to elucidate causal relationships between the mental disorder and the criminal act, and found indications of causal relationships in 17 of 75 cases of serious crimes, and 25 of 26 cases of homicide (Skeie & Rasmussen, 2015). The authors interpreted the results as an indication that Norwegian penal theory "*may result in impunity in a considerable amount of rulings where the illness of the accused apparently had no effect on the acts committed*" (Skeie & Rasmussen, 2015, p. 327). However, their results, their interpretation, as well as its significance rely on two premises: First, it remains to be explained why mental illness *as a causal factor* should excuse wrongdoing (Moore, 2014). Second is the assumption that mental illness can be distinguished as a cause, and that such causal relationships can be assessed at all, which might not be the

case (Moore, 2009, 2012, 2014; NOU 2014:10; Thorvik, 2008). In fact, it might not even be clear why even the mere presence of mental illness should play *any* significant role in excusing wrongdoing in the first place (Waller, 2011, 2014). Keeping that in mind, one might find it remarkable that, with the exception of Norway, most jurisdictions accept causality as a coherent criterion of moral responsibility.

1.4.1 Why is mental illness an exculpable cause of crime?

Why then, is it so often implicit that wrongdoing as a *result* of mental illness should be excused? Moore (2014) assumes the following reasoning: Legal punishment relies on moral responsibility, while mental illness can be regarded as a disability. An intuitive link is that no one should be held morally responsible for having a disability. Thus, there may be reasons for inferring causality in tests for legal sanity. However, given that it is very likely that every aspect of human nature can be reduced to some set of causal factors, why should wrongdoing caused by mental illness be *more* relevant as an excuse from moral responsibility than any other cause?

It is assumed that mental illness is a disability that inhibits our freedom to act as expected and culturally sanctioned. Intuitively, not wanting to act according to expectations may be more 'blameworthy' than not being able to. However, this distinction may not be as clear-cut as it seems. As Schopenhauer notes: "Man does at all times only what he wills, and yet he does this necessarily. But this is because he already *is* what he wills." (Schopenhauer & Kolenda, 2012). Who could disagree? It is hard to imagine how anything else is but an instance of dualism. So why are we thought to be more responsible for any other causal factor, including what may constitute our beliefs, intentions and wills, while *mental illness* as a cause is exculpatory?

Moore (2014) lists a set of popular maneuvers when confronting this question:

First, one could claim that mental illness as *a stronger* cause of behavior than causes *not* related to illness makes illness exculpatory. But this point is irrelevant, he argues: Yes, mental illness, however you define it, could be a stronger cause than other factors (however they are defined). But this would not matter because no less than 100 % of a set of causes leading up to a given behavior are sufficient to produce the behavior in question. On the other hand, one

could challenge this argument with the following example: Consider a university professor who loathes grading papers, yet merely remembering that it is part of his job description is sufficient for him to do that task. He knows it is not necessary to raise his salary or threaten cruel punishments for him to continue doing the task. Thus, he could argue that these are redundant causes that may overdetermine the choice in question. However, this argument may have three problems: First, the probability of continuing a certain behavioral pattern in the future is not what is being assessed in the court of law. The question is whether the perpetrator could have refrained from committing a *specific* act of wrongdoing at a certain point in time.

Secondly, the argument may rely on the premise that humans have some kind of privileged access to, and ability to predict and explain their thought processes and behavior, which might not be the case (Asch, 1956; Delgado, 1969; Haney, Banks, & Zimbardo, 1973; Johansson, Hall, Sikström, & Olsson, 2005; Milgram & Van den Haag, 1978; Nisbett & Wilson, 1977; Wilson & Dunn, 2004).

Third, it may require embracing a functionalist philosophy of mind in a certain kind of way (Kim, 2000; Ross & Spurrett, 2004; Ryle, 1949): High order abstractions, psychological constructs and mental entities, defined by a functional role instead of the physical constitution on which they supervene, are reified, and intuitively assumed to have a specific causal role. Functionalist psychological constructs may certainly be convenient as well as empirically valuable, but our intuitions about them as causal entities often fail. For instance, the assumption that increasing the incentive for a certain behavior is an additional, redundant cause also relies on a *second* premise regarding the nature of its causal influence. Possibly, to the contrary of what many would expect, studies have shown that participants given $1 for completing a tedious task rated it more fun and enjoyable than those who were given 20$ dollars (Bem, 1967; Festinger & Carlsmith, 1959).

This can be interpreted in many ways, for instance: Changing ones evaluation of the task from tedious to more enjoyable might resolve an unpleasant *cognitive dissonance* elicited by receiving insufficient incentives (Festinger & Carlsmith, 1959). Another explanation may be that merely by observing first- or third party behavior, participants might retroactively justify

it either by the nature of the task itself, or the incentive, dependent on its size (Bem, 1967). Furthermore, making threats may indeed induce compliance in some cases, but would threatening with cruel punishments really *increase* the likelihood of a professor doing his tasks? After all, who would want to continue working for an employer who suddenly threatens with cruel punishments? Actually, it should not be surprising that styles of leadership with low consideration for their employees, such as lack of warmth, respect and trust, seem to correlate with higher employee turnover rates (Fleishman & Harris, 1962). Clearly, our intuitions about high-level abstractions can fail, but *how* they fail is irrelevant in this case, the point is that they often *do*. It follows that making claims about the causal influence and redundancy of high order abstractions and psychological constructs may not be without its problems.

A second maneuver listed by Moore is the notion that the mentally ill are simply *less free from causation* (2014). In other words, mental illness is an exculpatory cause because sane people have more of what could be called ***contra-causal*** free will. Even contemporary philosophers propose such explanations (beliefs): Campbell (1957) proposed this kind of defense to save free will and moral responsibility, by suggesting a special power of choices in that nothing causes them, other than the choice itself. Furthermore, Sartre (1989) coined the neologism *being-for-itself,* which he argued is not subject to natural causation. Also, Glueck (1962) proposed a freedom-determinism-graph for adults to represent their capacity for free choice between 0-100%. He then asserted that the 'feebleminded' may be endowed with only 10% of intelligent freechoosing, and that 90% of their conscious purposive choice is blocked and determined. Psychotics could have somewhere between 10% and 40% of free choice, the sociopath between 30% and 45% and the 'genius' somewhere between 70% to 90% of free choice. These percentages may reflect the perpetrators *"capacity to intervene at the outset, and at various stages in initiating or modifying a causative sequence*" (Glueck, 1962, p. 16). However, in my view, this capacity must also be a result of antecedent causes. Besides: How can anyone 'modify' the causes of their own behavior, when any act of 'modifying' itself arguably would be part of their behavioral repertoire?

A similar *third* kind of maneuver is the notion that mental illness is a strong, sufficient or direct cause of wrongdoing, while 'normal behavior" only predisposes or is weakly causal on human behavior. Normal behavior would then consist of traits such as greediness or

melancholy, and only predispose, never determine choices. The assumption that strong causes leave less room for 'free' choice could be regarded as an example of dualism: Not knowing the full causal story is confused with the idea that there is nothing to be known (Moore, 2014).

A *fourth* kind of maneuver is an assertion that *because* we are less ignorant about how mental illness causes our behavior, the behavior can be excused. Yet this only sheds light on why we might be *tempted* to excuse apparent pathological behavior, not why we *ought to*. Indeed, in absence of mitigating information, an apparent lack of effort evokes more anger, while an apparent lack of *ability* evokes more sympathy (Weiner, 1993). Still, why should third party knowledge (or lack of knowledge) of the causes prior to someone's behavior have any bearing on his or her moral responsibility?

A *fifth* kind of maneuver is what can be called **selective determinism** (Moore, 2014): Even though our behavior must be fully caused by a set of factors, we only consider a few of them, and to a various degree assume their strength. For instance, whether to regard cultural ethnicity or poverty a possible causes for theft, or pathological personality traits or delusional beliefs as 'stronger' causes of wrongdoing than religious ideation (O'Connor & Vandenberg, 2005). However, this only signifies that it is *ideology*, not mental illness that dictates moral responsibility: The deciding factor is simply a cultural preference.

So if *excuses* based on the causality of mental illness seem to extend to *universal excuse* by *all* causal factors, should *everyone* be excused from their crime because no one can be morally responsible? Bruce Waller (2011, 2014) labels such counter arguments as the straw man of **excuse-extensionism**, which may be the product of being so deeply immersed *within* the system of moral responsibility that one cannot even imagine that there may be any alternative to it. This kind of immersion might also be what leads Daniel Dennett to warn about the consequences of uncovering the causes of our behavior (Dennett, 1984, 2004). Thus, he coins the phrase *creeping exculpation*, and proposes the peculiar solution that we should draw a limit to deep, careful scientific inquiry about the causes behind human behavior.

So why is it the case that not being *morally responsible* (to not deserve punishment) leads to the conclusion that a wrongdoing should be *excused?* If universal excuse is believed to be the result of knowing the causes behind human behavior, no wonder many are fearful of uncovering them. Nevertheless; so far there seems to be no reason to favor mental illness as an *excusing* cause for wrongdoing, or to assume that such casual assessments are even possible.

1.4.2 Inferring the causes of human behavior

Still, if one were to entertain the thought that mental illness somehow could or should be regarded as causes for wrongdoing, how could we infer such causal links? As Moore (2009) argues, the usage of the word causation in legal theory may have little in common with causation as we know it in a naturalist framework. Thorvik (2008) notes that causation may not be directly observed; only postulated. Moreover, the uncertainty of such postulations may increase depending on the immediacy of the given factors. Thus, the etiology of human behavior, given the endless complexity of the influential factors within biological systems as well as their environments, becomes unclear.

So how can causes of criminal behavior assessed and established? In one study, such assessments were made mainly through accounts given by the offenders, and rating their stated motives to estimate whether they were misrepresented, whether they were pathological, and finally whether both the rater and the offender considered the motive as 'directly causal' (Taylor, 1985). Indeed, I find it difficult to imagine other possible approaches to postulate such causal links. However, if we accept the premise of that it is feasible to make such causal assumptions for specific wrongdoings, it nonetheless seems possible to identify three sources of error in the study:

First of all, the validity of the assessments rests on a premise of having made a meaningful distinction between pathological and benign belief. Second, it rests on the prospect of making valid and reliable assessments of honesty. Third, the approach presumes that people, regardless of whether they are criminal, mentally ill, or delusional, have some privileged access to their previous or current mental processes, and that they can retrospectively retrieve and make valid claims about the significance of these processes in regards to their crime.

Contrary to common views, we are quite unaware about the actual psychological events that guide our choices (Nisbett & Wilson, 1977). A meta-analysis found an average correlation between intention and behavior to be .47 (Hausenblas, Carron, & Mack, 1997), but that does not account for the origin of the intention, or its causal relevance to the given behavior. In other words, even if there seems to be a moderate correlation between an intention and it its respective behavior, we must not forget that correlation does not necessarily imply causation.

For instance: Picture a bar where two guests are playing pool. It may be easy to infer a causal link when two billiard balls bump into one another, and the causal explanation *may* be sufficient *for its purpose*. On the other hand; postulating a cause for why the pool player picked up the cue in the first place may prove a lot more difficult. Of course, one could just ask the player, and he may state his own reasons such as: *"it was my turn"*. Other reasons he could give: "*It has been a long time since I played some pool, so I decided to have a go"*, "*my friend needed to go to the restroom, so I am playing his turn*", or "*we´re playing pool because the shuffle board was occupied*". Ask him the same question after a few beers, and he might reveal that he decided to stay at the bar because of a fight he just had with his spouse. None of the causes he gave are necessarily incompatible with each other. They may all simultaneously seem true from his subjective standpoint, and so they may all be answers to the question "Why did you pick up the cue?" Colloquially, it may be convenient to assume intentions as causes for behavior. Intentions may indeed correlate with their respective behavioral patterns, but, of course, correlation does not necessarily imply causation.

Nisbett and Wilson (1977) challenged the assumption that we can observe our cognitive processes by introspection, concluding that many of the verbal reports by introspection are merely retroactive confabulations. They provided several hypotheses that may account for the inaccuracy of introspection, for instance: Confusing mental content with mental process, a lack of disconfirming feedback to test explanations, various cultural influences, and the unconscious motivation to construct a coherent narrative (Wilson & Dunn, 2004). For instance, consider the observation made by Delgado (1969) when he evoked seemingly normal patterns of behavior such as head turning and slow body displacement, by means of electrical brain stimulation: This was repeated six times on two different days, and the patient always provided a reasonable explanation for behavior that was elicited. When asked *"What*

are you doing?", the patient would readily respond "*I am looking for my slippers"*, *"I heard a noise,"* or *"I am restless",* and so on (p. 116). The author found it difficult to ascertain whether the stimulation evoked the movement directly, which would mean that the explanations to some degree were confabulations, or indirectly by eliciting hallucinations which may have induced the patient to move.

Johansson et al. (2005) investigated participant's introspective abilities by asking them to make choices between face pairs based on attractiveness, and then covertly exchanged the photos, so that the outcome of the participant choice was the opposite of what was intended. The majority of the participants did not notice the change, yet often provided explanations for the outcome of their choice, for instance: *"I chose her because she had dark hair",* even though the person they had chosen was blond (Johansson et al., 2005, p. 118). This is clearly a sign of retroactive confabulation, an effect the authors named *choice blindness.*

Furthermore, psychological constructs and entities such as beliefs, attitudes, values and personality traits, can only be a part of the causal influences on our behavior. Consider the pool player, and the antecedent causes leading up to his behavior: Could it make a difference if he had slept well the prior night, or if he recently had a nutritious meal, increasing his energy level and his confidence as a pool player? For instance, Baumeister (2002) found that everyday acts of self-control such as delaying gratification, and rest/sleep depletes and replenishes our resources respectively, significantly influencing later choices and general functioning, and have coined the term *ego depletion*. Thus, the choices made at one point during the day may be influenced simply by what choices were made earlier that day, for instance, by deciding not to eat the chocolate cake in the fridge. Furthermore, Gesch, Hammond, Hampson, Eves, and Crowder (2002) conducted a placebo-controlled, double-blind study on adult prisoners, and found a 35.1% reduction of antisocial behavior as opposed to baseline, merely by adding dietary supplements to their meal. So then, simply adding dietary supplements prevented a significant amount of individual instances of violent, possibly punishable behavior.

Even subtle situational factors can have significant impact on moral behavior: Pärnamets et al. (2013) found that even while reasoning about a dichotomous choice on high-level moral

issues, the choices could be biased toward a target simply by gaze-contingent manipulation of the subject's immediate environment (terminating deliberation at the moment the participants gazed on a target option). Moreover, Isen and Levin (1972) conducted an experiment showing that simply finding a dime in a telephone booth influenced subsequent helpfulness (helping an actor pretending to drop her papers). Nearly all of those who recently found a dime stopped to help the actor, while those in the control group seldom helped. The subjects themselves did not associate their behavior with the independent manipulations. The authors interpreted the results as an indication that feeling good could lead to helping behavior.

If one were to accept a causal criterion commonly found in insanity tests based on the psychological principle, would it not be legitimate to regard finding the coin as *the* cause of the helping behavior? Certainly, the helping behavior was more moral than not helping, and uninformed observers would perhaps rate the helpers as more likeable. However, would it be fair to regard those who helped as more *deserving* of praise, considering that their helping behavior was almost entirely dependent on the independent variable, namely the mere absence or presence of a dime in their immediate environment? If a subtle situational manipulation was close to a perfect predictor of a specific subsequent act of kindness, where is the 'freedom' or moral responsibility in their morally superior choice, in that instance?

Milgram (1963) conducted a famous experiment studying obedience to authority: 49 males were recruited through advertisement, believing that they were participating in a study of memory and learning. They were fooled to think that they had been randomly assigned the role of "teacher", and were to subject the "learner" (an accomplice of the researchers) to electric shocks following wrong answers, to see the effects of punishment on learning. Presented with sample shocks of 45 volts on beforehand, to confirm the alleged authenticity of the generator, the participants presumably believed they were administering real shocks at 15-volt increments. Despite signs of emotional stress in the participants, such as an odd nervous laughter, and despite the fact that the "learner" at the 300 volt point started to pound on the walls, and subsequently stopped pounding as well as responding, 65% (26 out of 40) continued to the end under no pressure beyond simply by being told to. At this point, the participants presumably thought that they had administered *lethal* shocks of 450 volts. Milgram found that amongst respondents predicting the behavior of hypothetical participants

prior to the experiment, expressed considerable agreement on expected behavior: The estimates ranged from 0 percent to 3 percent, with mean of estimation of 1,2 percent to continue following orders to the end (Milgram, 1963).

This is a remarkable difference from the results of 65% of the participants in the actual experiment. Connecting this to the current state of our judicial system, could one claim that assessments of causation and moral responsibility of criminals at least *might* be unreliable? Would uttering the words "you could have done otherwise" to any of the 65% whom administered lethal shocks, be anything but absurd? Similarly; imagine an 18-year-old who commits premeditated murder, and that it is later revealed that his authoritarian father had been verbally coercive, telling him to do so: Would a jury not be tempted to assume that the perpetrator of course "could have done otherwise" by 'simply refusing to obey? Moreover, when presented with empirical evidence on obedience to authority, they might dispute: "Except that was an experimental situation, hardly generalizable". That might be the case, but then again, we must not forget that the participants in the experiment certainly did not know that. They would have done the same thing regardless of whether the shocks were nonexistent or *lethal*.

So, then if a jury, not familiar with Milgram's experiment, was presented with the superficial narrative of *any single one* of these participants, would they have much room left to express empathy towards the "teacher", considering that the "learner" seemed to have been tortured and killed? Bearing in mind that video evidence presented the defendant as follows; laughing whilst continuing to administer shocks of higher voltages, with audio of the learner pounding on the walls; would judgment be improved as to be more objective? Is it reasonable to suspect that attribution errors (Gilbert & Malone, 1995; Jones & Harris, 1967) could play a significant role in maintaining retributive systems of moral responsibility? Some might look back at Milgram's findings, and be tempted to argue that they can *no longer* be generalized. However, culture may surely change considerably over 50 years, yet not so much our human nature. Besides, according to a meta-analysis, replications show a remarkably consistent rate of obedience, often identical to Milgram's findings. Also, there was no significant correlation between the time when a study was conducted, and the rate of obedience, possibly serving as evidence against both time-effects, and enlightenment effects (Blass, 1999). Thus, it seems as if Milgram's findings may be just as applicable today as they were 50 years ago.

One thing we might gather from these findings is the following: Situational factors can have a strong influence on human behavior, and their effect, salience and strength is essentially a subjective, arguably unreliable supposition, especially if asserted without closer inspection (for instance, in the court of law). Consequently, it seems impossible to know the sufficiency of most, if not all of the causal assumptions regarding a *specific* act (whether it be an everyday choice, an act of kindness or a case of premeditated murder). In my view, it seems as if *necessary* causes are confused as to be *sufficient* causes, which may lead some to believe that mental illness as a *direct cause* of anything can make any sense at all. Mental illness needs an entire organism, and its environment, through which it could exert its influence. It follows that sufficient, direct causality by mental illness appears downright incoherent. This does not mean that it is not a *necessary* cause of wrongdoing. Then again, they *all* might have been. So if mental illness per se cannot be assessed as an exculpatory cause for criminal behavior, how can it be relevant for responsibility?

1.5 The medical principle

Moore (2014) argues that given the incoherency of insanity tests that are based on the psychological principle, 'craziness' should only be a *status excuse*. This means that the mentally ill simply lack the capacity for being moral agents the same way as children, non-human animals and inanimate objects. Thus, the medical principle must rely on biological and medical cues *per se* for deciding legal insanity, while inferring causal connections should be avoided. Accordingly, the nature of the act, its prior motivations and reasoning should be irrelevant for determining legal insanity. Only the mere presence of a psychiatric diagnosis, or a certain state of mind, is necessary and sufficient (Moore, 2014).

This makes the Norwegian example a somewhat purebred Kantian cognitivist approach (Thorvik, 2000). Norway may be the only country that currently makes use of such a pure medical principle (Moore, 2014; Syse, 2006, 2014). The main reasons for this is the assumption that a deviating psychological state affects "the whole personality", and that proving a certain state of mind is viewed a lot less challenging than proving its causal effects (NOU 2014:10 p. 48). For instance, a criminal act may seem to have reasonable motives, despite the fact that a clinical psychologist would label the reasoning pathological on closer

inspection.

The Norwegian penal code states that those who are "psychotic" are not to be punished, but it is important to note that the medical and legal understanding of the term "psychotic" are not supposed to be identical. The legal conception includes a varied set of mental disorders, with different types of etiology and quality. An important criterion in the legal conception is the state of mind at the moment of the act. Mental disruptions occurring before or after do not relieve from criminal responsibility (NOU 2014:10, pp. 49-51). The legal conception of psychosis considers the severity of psychotic symptoms while the crime was committed, whereas the corresponding medical diagnosis also could include individuals lacking the presence of major active symptoms.

Moore (2014) argues that the question of legal conception of insanity should be dependent on some *Factor X,* which must be evaluated on an individual basis, by means of a de-psychiatrized mix of several mediating factors that may be relevant, for instance, mental illness, substances, immediate situational circumstances, and personality traits. The *legal* conception of 'psychotic' is a variant of this. Moore's Factor X (and the medical principle) postulates a *threshold* for a minimum level of competence sufficient for some kind of *plateau* of moral responsibility (Waller, 2011, 2014). This is why Moore criticizes Norwegian insanity laws, and especially the psychiatric expert witnesses in the trial against Breivik for confusing the proxy (mental illness) with the thing being proxied (Factor X), by relying heavily on the criterions listed in ICD10 for paranoid schizophrenia, Moore argues that medical nosology is designed for an entirely different purpose. Thus, he argues that using criterions for a medical diagnosis in the court of law is simply an effort of assigning "guilt-by-lottery" (Moore, 2014). One may be left wondering whether Factor X, and whatever legal criterions necessary for it, could be subject to the same criticism (indeed, the test is to a large degree about who *deserves* the conditions of a prison sentence, not about who committed a wrongdoing in the first place).

All on this plateau are to be accredited some general competence sufficient for being morally responsible agents, similar to the unique human rationality postulated by Kant. By this model, criminals are regarded as morally responsible because of their alleged sufficient capacity to

appreciate what they *should* have done, and further it is often assumed that it follows that they *could* have done otherwise than what they actually did (Waller, 2014). *Above the threshold*, and beyond, any upbringing, traumas, current psychosocial stressors, or immediate, but seemingly subtle situational influences are ignored: Those on the plateau not only *ought* do otherwise; they also *can* (Waller, 2011, 2014). Furthermore, since criminal trials always take place *after* the crime was committed, *can do otherwise* turns to *could have done otherwise*. In addition to "psychotic", the legal conception of unconsciousness, and "disturbance of consciousness" may satisfy, or contribute to the threshold. That is not limited to comatose states, but includes what can be regarded as 'relative unconsciousness', which are states where motor activity is intact, and the individual is reacting to environmental stimuli, but lacks the ability to reason properly, and therefore *the usual self* (NOU 2014:10). Such states could be sleepwalking, hypnotic states, epileptic seizures or brain concussion. Proving the extent of such a state of altered consciousness has proven to be quite difficult. Moreover, those suffering from "severe mental retardation" are not criminally responsible according to Norwegian criminal law (NOU 2014:10).

The legal definition of severe mental retardation is also related to medical diagnostics. An Intelligence Quotient of somewhere around 55 and lower may qualify, but the threshold is also considered in light of factors such as personality traits and social function (NOU 2014:10, p. 52). Certain types psychopathology may have fallen outside of the definitions of "psychotic", "unconscious" or "mentally retarded" (NOU 2014:10) which is why some have proposed to expand on current theory (Rosenqvist, 2012). Lastly, the age of the perpetrator is considered. In Norway, the age of criminal responsibility is 15. The reason for this is the assumption that children are less able to "resist their will" than adults, because of their limited cognitive capacity (NOU 2014:10, p. 53), or because their free will has not yet been sufficiently developed (Syse, 2006).

Moore defends the medical principle while raising issues about the habit of giving special treatment to mental illness: "If Factor X is already a responsibility- eliminating or diminishing factor, independently of any exculpatory work done by mental disease itself, why does it matter how Factor X came to exist in a particular case?" (Moore, 2014, p. 17). So far, it does indeed appear as if the medical principle could have potential to be a more reliable measure

than any test based on the psychological principle.

However, if we analyze the claim that mental illness "affects the whole personality" (NOU 2014:10), what does it mean? Indeed, a person may commit a wrongdoing in relation to 'bizarre' beliefs about a demon, or in Breivik's case, right wing extremist convictions, and beliefs about the dangers of multiculturalism, as well as 'bizarre' beliefs about him being a member of something called 'Knights Templar' (Melle, 2013). Yet, why should having beliefs in such oddities lead us to conclude that people are not 'themselves' any more than those who hold less bizarre beliefs? Do 'bizarre' beliefs have an inherent property that limits the perpetrator's empathy, and capacity to weigh their value and importance against *rational* beliefs, such as the concern for the well-being of 77 other human beings and their next of kin? Is there really any evidence that the 'bizarreness' of his beliefs is one of the few-, or *the* factor that determines whether he can stop himself from acting of them? Is it not just as much his *selfishness* and raging *narcissism* of catering to his own ideological agenda, with complete disrespect for the opinions and lives of others that led him to commit his crimes? There may certainly be psychological benefits from expressing contempt for such bad character traits, as it may shape his character and as well the characters of others, but is there anything unique about these psychological constructs that makes someone more responsible for having them, or for their consequences.

Conclusively, might not a perpetrator's wrongdoing be better viewed as an outcome that arrived possibly *by means of any*, and as a *result of all* the past and current influential processes within the entirety of his *body* as well as the past and current influences of his *environment*, no matter how subtle? These influential factors may perhaps be assessed by psychologists and psychiatrists, but so far, the utility of such assessments certainly do not seem to have anything to do with the notion of moral responsibility and the retributive principles of our current penal theory.

1.5.1 Keeping our system of moral responsibility

The retributive basis of Norwegian law, assumes that perpetrators can be blamed for their actions, and thus *deserve* their punishment (NOU 2014:10, pp. 404-405). However, it is important to note that retributive basis of the Norwegian legal system does not *necessarily*

limit discussions about the *utility* of different kinds of punishment enforced upon the morally responsible. In fact, this has become a common, modern way of keeping the system of moral responsibility (Waller, 2014): The act of defending an incoherent retributive theory of punishment by means of applying consequentialist arguments regarding the extent and utility of various types of punishment, and of excuse. Thus, when it comes to quality of criminal care, Norway can become one of the leading nations (Waller, 2014), despite the fact that its judicial system is based on a false premise.

On the other hand, the notion that we must excuse the 'legally insane' from moral responsibility and thus from their wrongdoings entirely, can certainly have a peculiar results, perhaps especially in within what many may view as an international moral high ground; Norway. Breivik deliberately murdered 77 people, yet the state wanted to exculpate him, in addition to building his own hospital unit, which he never wanted, never used, and possibly did not even need. To address this is not moral philosophy. It is simply a common-sense observation of an obvious practical oddity, which in part may be the result of a widely accepted, but false axiom about human psychology, namely free will and its usual baggage of legitimizing moral responsibility. Additionally, there may be other beliefs and psychological phenomena that may add to the resilience of what Waller (2014) labels the *stubborn* system of moral responsibility.

So then, the utility of different kinds of sanctions can be discussed in relation to, for instance, criminal deterrence, compassion for the perpetrator, and maintaining social order (NOU 2014:10). Traditionally, arguments have largely focused on public and individual deterrence. For instance, both kinds of arguments can be used as a justification for exonerating the mentally ill: First, certain individuals might have little if any benefit from punishment, effective neither as deterrence, nor as rehabilitation. Typically, individual deterrence-arguments regard the effects of treatment: On one hand, on could claim that imprisonment is not proper treatment. On the other hand, it is believed that holding people responsible could actually be beneficial from a treatment perspective (NOU 2014:10, p. 405). However, Waller (2014) notes the large amount of evidence showing that punishment is not an optimal shaper of behavior, and that there are too many cases indicating that it is counter-productive.

Juth and Lorentzon (2010) claim we should remain 'neutral' or 'agnostic' to the metaphysical conception of free will, given the lack of scientific consensus regarding its existence. However, "I don´t know" is hardly the most suitable answer to the existence of an unproven entity, especially in light of prior knowledge. Any Bayesian should agree that given the overwhelming, and continually increasing amount of available evidence that sheds light on the constraints on human behavior, it should certainly not allow for a starting position that is "neutral" about the freedom of our will. This point becomes of great importance when taking in to account that such beliefs may have substantial implications for the well-being of *many* (Waller, 2011, 2014).

1.6 Legal insanity and confirmation bias

Arguably, even if there are cases in which legal insanity has seemed to be a useful construct, for instance when an offender's temporary delusions involved a conviction that his murder-victim was a demon from hell. However, such a conclusion might only ensue because the construct legal insanity has emerged within a cultural context in which the notion of moral responsibility and retribution is also accepted: "Retribution for wrongdoings is usually sensible, but sometimes retribution and harm and for harms sake is unjust." Hence, any apparent construct validity attributed to legal insanity might simply be proxied by confirmation bias: If retribution and harm for harm's sake is always unjust, then it is obvious that it is also sometimes unjust.

1.7 The question-begging tests for legal insanity

As mentioned earlier: criminal responsibility is ordinarily not determined solely by criteria for a medical diagnosis of mental disease, but by the medical or psychological principle for legal insanity, which limits the use of punishment (retribution). However, it appears that these considerations may be pointless, and the theoretical distinction between the medical principle and the psychological principle a false dilemma: By rejecting the incoherent criterion of causality, the medical principle may appear to have potential as a more reliable measure. Yet, no matter how reliable it might be, it is perhaps nonetheless simply to divide by zero, given that the target appears to be nonexistent. It follows that both principles seem to be incoherent. In addition, it follows that any divergence in criminal care that may result from making distinctions based on these assumptions is likely to be unfair.

However, by being immersed within the specifics of this possibly futile theoretical discussion, one might never be tempted to reflect upon, or to challenge its legitimacy. As crucial partakers in making distinctions between the deserving and the excused, psychologists and psychiatrists should arguably be sensitive to the following: Large amounts of theoretical work regarding legal insanity, as well as the application of these theories in the courtroom, might be a wasteful drain of valuable resources, as it may be nothing but a way to legitimize retribution by means of pseudoscientific inquiry into the possibility of an impossibility: Free will, and/or moral responsibility. As I have mentioned, making these distinctions does not necessarily limit discussing the utility of different kinds of sanctioning. Still, the extent and impact of such discussions, is in danger of being limited, attenuated, and deafened by the notion of just deserts. It also follows that the mandate given to psychiatrists and psychologists in criminal trials arguably is theoretically and ethically incompatible with their profession.

For instance: Is it really surprising that the expert medical witnesses in Breivik's trial tended to focus on entirely on medical criteria? Do medical professionals usually have any formal training and discernible skills in assessing some vague notion of legal insanity by assessing causality in systems of incredible complexity, and/or some question-begging threshold of moral responsibility? In fact, should anyone who professes in the treatment of illness, or who swore the Hippocratic oath, or for that matter, anyone at all, by any means, directly or indirectly partake in the process of ascribing any amount of harm for its own sake? Moreover, given the opposite conclusions by the expert witnesses in Breivik's trial, was there any benefit from their costly contributions demonstrably palpable?

Might there be some better way for psychology and psychiatry to contribute in the courtroom when it comes to criminal care, deterrence and prevention? Is it possible to construct a legal framework in which clinical psychologists and psychiatrists can contribute while professionally uninhibited by the unsubstantiated demands made by systems of moral responsibility? Fortunately, a unique example of an exception does exist, namely Swedish penal theory.

1.8 The Swedish system of criminal care

Influenced by positivist, preventive theories of penal law, Sweden is unique by rejecting the notion of moral responsibility (Juth & Lorentzon, 2010). According to their model, the mental

state of the perpetrator is not to be treated as a condition for criminal responsibility. Instead, it is only relevant in regards to extent of punishment / sanctioning. The model is not guilt-oriented, but need-oriented: Only future benefits should be considered as justifications for the sanctions that are proposed. The question following a conviction is not whether or not to place blame for the sake of retribution, but mainly to decide how to prevent such a thing from happening again. The resulting sanctions were to be fitted to the societal and individual needs. It is based on the assumption that a crime is always determined by a "defect" within perpetrator and his environment (Juth & Lorentzon, 2010). Thus, moral responsibility is regarded a pure metaphysical entity, incompatible with reality and is not to influence penal law (NOU 2014:10). It follows that people are expected to follow the law, and can be held responsible; not as *moral agents*, but as *legal subjects*. If someone did not act as expected, by committing a wrongdoing, there is no reason to claim that they could have done otherwise than what they actually did in that specific instance, regardless of how "competent" they are perceived to be.

Waller (2011, 2014) proposes the terms *role responsibility* and *take-charge responsibility* as a contrast to moral responsibility, arguing that the latter is not compatible with a naturalist world-view. To clarify this argument: Even if we cannot ultimately be blamed for our character and our actions, we may be able to monitor ourselves, learn by our faults and our fortunes, respond to criticism and praise, and *take* responsibility for our past- as well as our future actions. When an airline pilot makes a minor mistake, he could receive a warning. He can try to be more vigilant, but if he still cannot fulfill the pilot *role responsibilities*, he might have to step down from the job, even if he cannot ultimately be blamed for his failures. Similarly, even if a criminal ultimately cannot be blamed for his bad character, he might be sanctioned as a routine measure for individual and public deterrence, or if he is violent, he may have to be contained, which would mean stepping down from the role as a free member of society. Hopefully, the criminal also acknowledges his wrongdoing and *takes* responsibility for it, giving room for hope, especially if he also had, say, an internal locus of control (Rotter, 1966), and a strong belief in his own self-efficacy (Bandura, 1977). Moreover, these constructs are compatible with a naturalist world-view, and with the reasonable notion that our volition is fully constrained by antecedent factors. Thus, a criminal can indeed be more role / take-charge responsible for his actions than a rolling boulder or a roof avalanche striking a random pedestrian, but not any more *morally* responsible.

If one were to argue "because he is not morally responsible, punishing him is *unjust*" a response to that dispute is as follows: It is *indeed* unjust that some end up legally sanctioned to punishment. There seems to be no rational reason for why anyone *deserves* any harm. However, if we hold someone *morally* responsible, that might dissolve the cognitive dissonance that should result from punishing those whom are likely to already be having a hard time. This might partly be why beliefs in a just world is a common cognitive bias amongst us (Lerner, 1980), laymen and academics alike. Even if it has not necessarily always been the case, most of us today would probably agree that it is absurd to claim that an inanimate object, such as a rock, or an animal such as an alligator, a lion or even a non-human primate would ever *deserve* any punishment at all, no matter how much harm they may have inflicted. So then, what is it about human psychology that could make such a claim in the court of law any less absurd?

Swedish law incorporates the rule of "prison prohibition", which states that mentally ill are not to be imprisoned. They are to receive proper treatment for their mental defect, although in principle, all criminals are supposed to be sentenced to care: The administration running the prisons is called the "Criminal Care authority". However, the "care" of Swedish prisons leaves a lot to be desired, and they would be regarded as institutions for punishment and incarceration in the eyes of most people (Juth & Lorentzon, 2010). Nevertheless, even though the practical dissimilarities of the Swedish model appear to be somewhat subtle, compared to, say, the Norwegian model, it may still be an incredibly important theoretical distinction in principle. Some have argued that Sweden should reject its current theoretical model, so that they can finally be internationally compatible (NOU 2014:10). In my view, this would be a dreadful step in the wrong direction. It is astounding that anyone that easily could make such a suggestion, given its possible implications.

2.1 Defining free will

The definitions of free will considered so far all seem to share the idea that it means that an agent "could have done otherwise" as he was making a "free" choice. These theories are largely divided between libertarianism and compatibilism.

2.1.1 Libertarianism

Libertarian arguments in favor of free will rest on the premise of indeterminism, and presume that humans have a unique ability to break free from the laws of nature as an uncaused cause insofar the agent is making free choices. Libertarian accounts are commonly regarded as the "magical" explanations of free will (Waller, 2015).

Descartes' dualism, for instance, raises the idea that humans are endowed with a soul and a godlike perfect "will", and that they are morally responsible for immoral behavior insofar they fail to apply this divine faculty properly (Descartes & Cottingham, 2013). Although this is might be an extremely short and unfair summary of Descartes thoughts about free will and moral responsibility, I nonetheless find it difficult to imagine how one could test these kind of claims. Hence, one might wonder whether we should give the Cartesian view of free will the benefit of the doubt.

Kantian rationalism is in some ways similar to the Cartesian view. Kant assumes the reality of a "pure will" that potentially can solely determine human choice "unaffected" by other impulses. Like Descartes, Kant believed that free will was limited to humans because of their uniquely human rationality. Human volition becomes free from the deterministic world, according to Kant, precisely by thinking of itself as free. As Kant writes:

> The will itself, strictly speaking, has no determining ground; insofar as it can determine choice, it is instead practical reason itself. Insofar as reason can determine the faculty of desire as such, not only choice but also mere wish can be included under the will. That choice which can be determined by pure reason is called free choice (Kant, 1996, p. 13).

I find Kant's arguments problematic for many reasons. First, he claims that an entity, namely the "will", can serve as an uncaused cause, which is arguably an extraordinary, dualist, claim. Second, it seems difficult to understand what "will" could be and how could we possibly observe this purified entity "will", and its influence? It appears, then, that this definition of

free will seems, at least in this case, to be untestable and thus unscientific.

It appears that that ideas about free will and retributivism tend to rest on the possibility of indeterminism (NOU 2014:10; Syse, 2014; Thorvik, 2000), but in every case they seem to leave out *how*. More recently, free will has been tied to **quantum indeterminacy** (Kane, 1998). However, even if the word "*quantum"* might seem appealing as part of a 'modern' libertarian conception for free will, that certainly does not need make it more coherent, explanatory, or useful. Like "consciousness", one might say that the nature of quantum phenomena appears quite mysterious to us. However, while it is indeed true that quantum physics is quite mysterious and poorly understood by most, if not all, of us, this should not permit us to trust some intuitive feeling that there is a connection between quantum physics and consciousness. In my view, such accounts for free will are just additional instances of 'the God of the gap', or more precisely, "quantum of the gap".

In addition, even if it was in fact true that quantum events are necessary for any type of conscious choice, I find it quite incomprehensible how it could possibly relate to current psychological terminology, or even to legal terminology (see also Latorre, 2013). If the quantum world is that which allows for free will and moral responsibility, does that mean that quantum physics should become part of the curriculum for anyone working in forensic psychiatry? Finally, if quantum physics enabled human free will and moral responsibility, why would it not do the same thing for other animal species? After all, are they not living in the same universe? Thus, I find reason to conclude that "quantum will"-definitions for free will are untestable and unscientific.

2.1.2 Compatibilism

Compatibilist positions for free will tries to unify free will and moral responsibility with determinism. In other words, compatibilism attempts to be compatible with determinism. Bruce Waller notes, however, that the there is an impressively large diversity and variety of different compatibilist accounts, and that they tend to be incompatible with each other (Waller, 2011, 2014).

First, compatibilist tend to defend free will and moral responsibility with **normative arguments:** This is when authors argue that we should maintain our current understanding of, and belief in-, free will and moral responsibility, regardless whether those beliefs are coherent or not.

This is for instance apparent when compatibilist Daniel Dennett asserts that we must be wary of looking to carefully at the details provides by science about the factors that constrain our behavior, because *"if we lose our view of ourselves as we gain in scientific objectivity, what will happen to love and gratitude (and hate and resentment)?"* (Dennett, 1984, pp. 13-14). Dennett fears "the boojum of creeping exculpation" (Dennett, 2004): the idea that scientific discovery shrinks the free agent to nothing, leaving us with no choice but to regard all crime as pathological, and thus must excuse all crimes. This is an instance of *excuse extentionism (Waller, 2011, 2014)*, which arguably derives from the failure to think outside the moral framework. This common dilemma is dramatized in Dilbert (Figure 1).

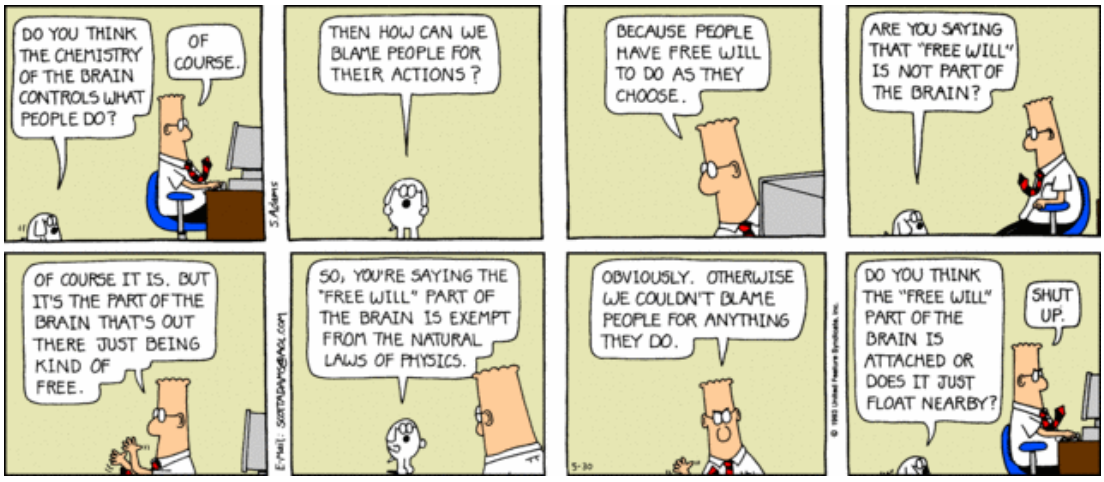

Figure 1. An example of excuse extentionism (Adams, 1993). Retrieved from http://dilbert.com/strip/1993-05-30

Compatibilist Harry Frankfurt proposed a **hierarchical model of free will** (Frankfurt, 1971). According to this theory, it is the strength of approval of one's own actions is that which make you responsible for them. Frankfurt distinguishes between first order desires and second order desires. First order desires might be simple desires, such as the desire to feed one's heroin addiction. Second order desires might then be the desire to live in sobriety despite ones addiction. These second order desires, according to Frankfurt, are "essential to being a person" (Frankfurt, 1971, p. 10). Agents who do not have, or do not care about second order

desires, Frankfurt labels “wantons”, which would be all nonhuman animals and very young children, for instance. Insofar there are conflicts between a wanton’s desires; these conflicts are irrational, according to Frankfurt, like when a savage creature is teared between immediate cravings for mating and feeding. It follows from Frankfurt’s definition of free will, that as long as there is no conflict between the first and second order desires, for instance if an addict reflexively approves of his own drug use, then he is choosing freely.

Frankfurt may not have provided any evidence to the claim that second order desires are a uniquely human feature. Still, a favorable feature of Frankfurt’s hierarchical theory of free will is that it seems, at least to a certain degree, potentially testable, scientific and compatible with determinism (or deems determinism an irrelevant aspect). For instance, it might not be so hard to imagine that there are people who do not intend or want to change their unhealthy drug habit, while there are other users of the same drug who despise their habit, as described in the Transtheoretical Model of Health Behavior Change (TTM) (Prochaska & Velicer, 1997).

There are crucial problems with Frankfurt’s hierarchical model, however. Consider a willing addict, for instance, who reports no conflict between his first- and second order desires. Are we right to say that he is necessarily acting *more* freely than those who do experience a conflict? According to the TTM, the willing addict seems fit the description of those who linger in the precontemplation stage of behavior change:

> People may be in this stage because they are uninformed or underinformed about the consequences of their behavior. Or they may have tried to change a number of times and become demoralized about their abilities to change. (Prochaska & Velicer, 1997, p. 39)

It seems to me that the willing addict might be willing because he has had a *lack* of opportunity. Perhaps he simply is in a stage of learned helplessness (Abramson, Seligman, & Teasdale, 1978), and adapted to this state by embracing his unhealthy habit. In other words,

perhaps the willing addict is rather *less* free than the unwilling addict is. So then, even though Frankfurt's account could provide important insights regarding the *feeling* of freedom, it might at best be an incomplete account of free will.

Some accounts of free will and moral responsibility can be referred to as **plateau models**. Common tests for legal insanity, for instance, seems to assume a plateau for moral responsibility (see section 1.4 – 1.5). The general assumption is that most people are presumed to be more or less equally "competent" to make moral choices, and that any differences are usually insignificant enough to be ignored. Hence, all on this plateau can be held morally responsible. While libertarian accounts for free will tend to share the idea of this plateau, the position is also held by compatibilists, such as Daniel Dennett:

> Moral development is not a race at all, with a single winner and everyone else ranked behind, but a process that apparently brings people sooner or later to a sort of plateau of development—not unlike the process of learning your native language for instance. Some people reach the plateau swiftly and easily, while others need compensatory effort to overcome initial disadvantages in one way or another (Dennett, 1984, p. 96).

Henceforth, one will blame the consequences of morally relevant choices on the agent. This is because if the actor was above the plateau, the idea is that he probably was capable of "doing otherwise" than he did (NOU 2014:10). As I have argued, this is an incoherent assertion, given that you cannot possibly test whether someone could have done otherwise than they actually did. In addition, plateau theories beg the question regarding where to place the plateau, and how to define it, and how to identify it on a case-by-case basis (see section 1.4-1.7).

Plateau theorists seem to believe that even if you are born with an advantage or a disadvantage-, whether it is socioeconomically, genetically, or otherwise-, most disparities will purportedly even out, and everyone can have their moment, at least if they want it enough. This notion seems to conflict with empirical evidence, however: Studies show that

success tends to promote more success, while failure increases the likelihood of more failure (Barash & Lipton, 2011; Demerouti, Bakker, Nachreiner, & Schaufeli, 2001), possibly via learned helplessness (Abramson et al., 1978), altered beliefs about self-efficacy (Bandura, 1977), or impacts on the amount of personal resources (Demerouti et al., 2001).

2.1.3 Closet libertarianism

It quite often seems as if compatibilists accounts in defense of moral responsibility turn out as libertarian accounts. In fact, in a chapter named "Giving Libertarians What They Want", (Dennett, 1981) describes the possibility of a "consideration generator" – a uniquely human unsubstantiated mental capacity which would give an indeterminate output, and that this indeterminacy is somehow responsive to the type of decision that is made. In my view, however, it might be sufficient to refer to this entity as "the brain". Dennett deliberates whether indeterminate processes could be interpreted as "physically determined, but random-, or patternless processes" instead of the metaphysical understanding of indeterminism. In my opinion, however, I find such discussions quite uninteresting, and fruitless. For how can this metaphysical discussion possibly help us understand-, or define a scientifically testable, legally applicable definition of free will and responsibility?

2.2 Free will without moral responsibility: Restorative free will

Bruce Waller (2011, 2014) challenges the usual conception of free will and moral responsibility, and presents an account of free will that, in my view, seems to have the most potential of being scientifically sound, testable, as well as useful. He defines his free will as follows: "*the capacity to effectively explore alternative paths in response to a combination of environmental contingencies and internal motives...*" *(Waller, 2015, p. 1)*

Waller calls his theory of free will **restorative free will** partly because it "restores free will to many species that have been denied free will under accounts contrived by humans to claim exclusive rights to free will." (p. 1). Waller separates Restorative free will from traditional accounts:

> Free will is not the quality that makes us godlike, nor is it the special gift of God to his last and favorite creation. Free will is not unique to humans, it does not require a high level of reflective rationality, it does not support moral

> responsibility, and it does not involve any special causa sui powers. […] On the restorative view, the free will powers enjoyed by human animals are basically similar to those of other animals, and free will—including the human variety— is better understood by examining the common elements of free will shared by many species, rather than focusing on the distinctive *enhancements* of free will that are unique to humans (Waller, 2015, p. 1).

The problems with other accounts of free will, Waller argues, is that although there are probably distinct features in human decision-making, our understanding becomes distorted when we try to focus solely on such distinctions, and treat them as the *essentials* of what we think is free will. Waller rather wants to understand free will like an evolutionary biologist, by tracing it back to the evolutionary processes that shaped free will for whatever value it might have had for survival of the species:

> If we wish to understand the human heart, first we start with the study of hearts in many species, and from that study we discover the basic structure and function of the heart; then we can go further, and study what is distinctive about the human heart. The same process is our best approach to free will: study the common basic function of free will in many species; and then study how distinctively human free will fulfills that function, and any special human adaptations to enhance the functioning of free will. (Waller, 2015, p. 3)

Moreover, Waller notes that as long as we regard free will as a uniquely human capacity that bears no relation to the behavior of nonhuman species, our theories have been severely limited empirically. Hence, our theories are not grounded in solid empirical understanding, leading to imagined, unsubstantiated, and yet, often extraordinary and miraculous claims. Perhaps one might argue that human decision-making is *remarkably* unique because of human language and verbal reasoning capabilities, and that it therefore should treated as "something else". However, as Waller argues:

> ..when we try to make reason into a special definitive power of free will—a power that makes free will a uniquely human capacity—the result is a distorted picture of animal free will (together with an exaggerated picture of human reason). The role of reason in human free will is best understood in the larger framework of animal free will generally. Reason enlarges our realm of effective exploration, as does the eagle's capacity to soar, the wolf's olfactory powers, and the hawk's keen eyesight. In that context, human reason is a valuable enhancement of free will rather than a defining feature. (Waller, 2015, p. 164)

Henceforth, phenomena like learned helplessness (Abramson et al., 1978), self-efficacy (Bandura, 1977), locus of control (Rotter, 1966), behavior variability in order to cope with complex contingencies (Kavanau, 1967), ego depletion (Baumeister, 2002), the affective value of perceiving choice opportunity (Leotti & Delgado, 2011), for instance, might all somehow serve as important insights for a sound conception of free will.

A unique feature of Waller's Restorative free will is that it sheds the usual ties with moral responsibility. Waller does not attempt to justify retributivism, but rather expands on what a sound definition of free will could be within a scientific framework, and explores how his conception could be relevant in our justice system. In final, it seems to me that Wallers "Restorative free will" is the only eligible candidate for a scientifically sound description of the term (see Table 1 for a summary).

Table 1. Summary of common accounts of free will, and their apparent features

Model	**Metaphysics**	**Free will is human exclusive**	**Moral responsibility**	**Testable / scientific**	**Crucial shortcoming(s)**
Cartesian	Libertarianism	yes	Yes	No	Dualism, Supernatural claims about the divine
Kantian	Libertarianism	yes	Yes	No	Vague, obscure, circular, relies on indeterminism (See 1.x)
Quantum will (Kane)	Libertarianism	Yes	Yes	No	Unintelligible, untestable
Normative	Irrelevant	yes	Yes	No	Not a definition
Hierarchical (Frankfurt)	Compatibilism	yes	Yes	Yes	(counterintuitive, incomplete)
Plateau (Moore, Dennett)	Compatibilism or libertarianism	Yes	Yes	No	Circular, question-begging
Restorative (Waller)	Irrelevant (naturalistic)	No	No	Yes	A revision of the common conception of free will

3.1 Implications

In this text I have argued that there are crucial problems with the free will axiom, even though it is currently a fundamental principle on which our retributivist justice system seems to rest (NOU 2014:10; Syse, 2012, 2014). More precisely, it appears that, although there might be scientifically sound definitions of free will available, the common understanding of the concept is incoherent. So how might our society respond if it were to reject the common understanding of free will? Some have proposed that it may lead to unwanted consequences such as attitude of resignation, harmful defeatism, or irresponsible behavior.

3.1.2 Bad effects

Vohs and Schooler (2008) found that reading arguments against the existence of free will increased participant's likelihood of cheating on a subsequent exam. Thus, the authors suggest that rejecting the common understanding of free will could lead to bad effects. There are several problems with their findings, however.

First, it has been noted by Nahmias (2011) that participants in this study are primed with one-sided descriptions of determinism, such as the notion that every human action is determined by the environment, and by genetic makeup, while any significance of conscious deliberation and reasoning is left out of the prime. Nahmias argues that this might lead participants to read the claims to mean that such conscious processes are *by-passed* by the deterministic processes in the brain. Perhaps that could be a likely interpretation, for participants who initially believes in libertarian free will, or that human reasoning and deliberation are not a part of the physical world (Cartesian dualism). In fact, it appears that controlling for such libertarian and dualist beliefs may make the subsequent bad behavior disappear (Nahmias, Morris, Nadelhoffer, & Turner, 2006).

Second, it has been found that weaker beliefs in free will leads people to endorse less retributive (backwards-looking) attitudes towards punishment of criminals, while consequentialist attitudes were unaffected (Shariff et al., 2014). In other words, to abandon free will could result in a positive societal shift that maintain the motivation to use punishment for any beneficial effects it might have, while decreasing motivation towards using it to inflict unnecessary harm.

Third, there is arguably a crucial difference between abandoning free will as a scientific term, on one hand, and explicitly preaching to all citizens the notion that "free will does not exist", on the other (Nahmias, 2011). Hence formally rejecting free will as a scientific concept does not necessarily concern the daily lives of neither the general public, nor a perpetrator until *after* someone commits a crime, perhaps by a shorter, less confusing, and more efficient trial procedure, in a society with perhaps a different, more reasonable set of laws.

Fourth, similar to the difference between the colloquial (non-technical) and scientific understanding of the term "theory", there can be a difference between the colloquial and

technical use of "free will". The claim "I have a theory about who took my sandwich" could perhaps colloquially mean; "I think I know who took my sandwich", and *not* "I have an empirically well-substantiated explanation for who took my sandwich". Similarly, the claim "he did it by his own free will" could mean "he wanted to do what he did", or "was not forced to do what he did". In addition, given that Breivik planned and carried out his acts on his own, with no gun pointed at his head, and with no subsequent remorse, the remark "Breivik killed 77 persons by his own free will" is, in my view, indeed appropriate in colloquial conversation. However, the legal question about Breivik's free will and moral responsibility seemed to concern a different understanding of the concept: Namely, whether or not he "could have done otherwise" and/or whether he *deserved* punishment for it (NOU 2014:10).

Finally, as noted earlier, whether the common conception of free will is coherent, is a scientific issue, not a political one. Indeed, after hearing about Darwin's theory, the Bishop of Worcester's wife reportedly exclaimed: "Dear me, let us hope it is not true. But if it is true, let us hope it does not become widely known." (Silk, 2016, p. 176). This is a sentiment that seems remarkably similar to how "proponents" of the free will-construct appear to argue today: An incorrect understanding of human nature is stubbornly maintained in fear of losing morality (Waller, 2014). If it were shown that beliefs in the Christian god, intelligent design, and the Ten Commandments somehow promoted peaceful behavior, should this mean that psychologists, politicians, and lawyers respectively should treat pseudoscience, theocracy, and biblical law as if it were coherent? It seems quite clear to me that they should not.

3.1.3 Excuse extentionism and funishment

Another common fear in regards to rejecting free will and moral responsibility is the notion that we would have to excuse every single perpetrator, since no one can be blamed for their actions (Dennett, 1984, 2004). This is what Waller (2014) calls **excuse extentionism**. Saul Smilansky also expresses this fear, arguing that "punishment" would have to be abolished if our justice system rejects free will, and that we would have to commit to "**funishment**" instead (Smilansky, 2011). Funishment would resemble punishment by the fact that criminals would be incarcerated securely to protect society. However, if we really were to believe that one *deserves* the suffering of incarceration, they need to be *compensated* for that, he argues. Hence the institutional facilities of funishment would have to be "as delightful as possible", and "resemble five-star hotels, where the residents are given every opportunity" (Smilansky,

2011, p. 355) and leeway to enjoy life. We could not have spared any effort or expense, according to Smilansky, in order to make sure that the residents suffer minimally from incarceration.

Indeed, similar to Smilansky's conception of "funishment", it appears that incarceration of criminals by compulsory mental health care must 'exert no limitations to freedom other than necessary for the protection of others' as per Norwegian legal theory (NOU 2014:10, p. 33). A legal insanity-verdict for Breivik, which was considered as highly likely prior to the second psychiatric evaluation (Syse, 2014), would have led to an estimated cost of incarceration somewhere between 30 000 and 50 000 NOK per day, partly in order to substitute door-locks with designated 24 hour triple staffing (Steen, 2012). I imagine that these costs could have financed not only a lifelong 5 star hotel room stay - it would perhaps be enough to finance dozen stays. Granted it will be more costly than a hotel-room to contain someone like Breivik. However, the absurdity lies in the fact that these costs easily could have been quadrupled, possibly for no tangible benefit whatsoever, except to serve our retributive principles.

3.1.4 Resolving the grey zone of legal insanity

Another problem that seems to derive from our current retributivist legal theory, is the grey zone that emerges as a result of the legal insanity-construct. Surely, everyone were confident that Breivik had to be contained no matter whether he had been deemed legally insane or not, which is why it might have been easy to forget that a legal insanity verdict in fact is a "not guilty"-verdict. In other words, it would mean that he was formally excused entirely for his crimes. As mentioned, the conditions for mandatory psychiatric treatment are remarkably high (NOU 2014:10); we cannot isolate a legally insane perpetrator if these criteria are not met sufficiently. In these cases, the legally insane perpetrators, whom otherwise would have been imprisoned and isolated from society, are able to leave the courtroom as free members of society.

One might assume that such instances concern only 'bothersome' perpetrators that commit minor offences, such as grocery theft, or public disturbances. In that case, one would be wrong: An investigative report by NRK **revealed** that between 2002 and 2010, as many as 135 perpetrators had been excused by legal insanity and condemned to mandatory psychiatric

treatment (Sanvik & Nilssen, 2013). However, the same report revealed that between 2002 and 2012, twenty-two violent offenders had been excused by legal insanity, but were also allowed to walk free without any sanction at all because they did not satisfy the strict conditions required for compulsory mental health care. Two of these individuals had committed homicides, four had attempted murder, nine were violent offenders, six cases of sexual abuse (including repeat offences against children), and finally, one was an arsonist, with a known history of at least 10 fires started).

Hence, it appears that 15% of legally insane wrongdoers of murder, sexual abuse, and severe violence during this period were excused and immediately released to the public, which arguably is not an insignificant number. When we view these cases in the context of our practice of punishing and incarcerating non-violent recreational drug users, for instance, we might start to wonder whether we are currently distributing our resources and our moral judgments wisely.

3.2 Justice without retribution

So how can we maintain a justice-system that is compatible with free will-skepticism? If no one deserves punishment, then how can we allow ourselves to incarcerate dangerous criminals?

3.2.1 Quarantine Model

Derk Pereboom proposed a legal framework, which he claims to be compatible with free will skepticism. In order to provide an answer to how we can justify sanctioning criminals, Pereboom draws an analogy with the practice of quarantining carriers of infectious diseases:

> The free will skeptic claims that criminals are not morally responsible for their actions in the basic desert sense. Plainly, many carriers of dangerous diseases are not responsible in this or in any sense for having contracted these diseases. We generally agree that it is sometimes permissible to quarantine them nevertheless. But then, even if a dangerous criminal is not morally responsible for his crimes in

> the basic desert sense (perhaps because no one is ever in this way morally responsible) it could be as legitimate to preventatively detain him as to quarantine the non-responsible carrier of a serious communicable disease. (Pereboom, 2014, p. 156)

Of course, this places some new constraints and requirements on how we can justify incarceration of criminals. It might perhaps be sensible to try to treat and rehabilitate someone whom we have chosen to quarantine, insofar we can. Nonetheless, if the quarantined individual carries an extremely dangerous and infectious disease, like smallpox, to which the general population is unvaccinated and vulnerable, then it would perhaps be a good idea to keep the individual incarcerated as long as he remains a threat. This might mean isolating the individual for the rest of his life, as long as the likely alternative was an outbreak of the disease. Nevertheless, we would not blame this person for contracting his disease. Hence, we would have to be apologetic about taking away his freedom, and maintain a concern for his well-being.

3.2.2 The Public Health-Quarantine Model

A problem with the Quarantine model is that it does not sufficiently consider broader societal factors (Caruso, 2016). Some questions that seem unanswered might for instance be; how infectious, and how harmful, does a disease have to be before it justifies quarantine? Likewise, how dangerous or violent does a criminal have to be before we can isolate him? What are the psychological consequences of being quarantined; how many will accept it as a necessary intervention for the common good, and how many would experience it as unjust and traumatic? How can we maintain their well-being while quarantined? When can we say that we have spent enough resources to compensate for their lost freedom?

The Public Health-Quarantine Model (Caruso, 2016), builds upon the Quarantine model (Pereboom, 2014), but also considers criminal justice analogous to how physicians consider medical ethics, and public health ethics. Public health ethics has four characteristics (Faden and Shebaya, 2015) that can be applied to the concept of public safety (as cited in Caruso, 2016):

1. the emphasis on the common good,

2. The focus on prevention,
3. The use of government intervention,
4. Consequentialism

First, like medical interventions, we should not consider legal interventions separated from the **common good**. For instance, there is no doubt that there are people who claim a need for medical treatment, and whom may be likely to benefit, yet who might not be prioritized in the health care system. This is because there will be some discrimination on a need-basis: How severe is the given health-problem, how painful is it, how costly is the intervention, and how effective has it proven itself to be? Moreover, politicians have to consider where resources are needed the most, and where will they benefit the most, in the broader perspective. There will always be treatments that are too expensive compared to the expected benefit. If we did not make such evaluations, healthcare would certainly be expensive, and the system would have recklessly demanded an unfair and amount of resources, in comparison with other public services.

Second, the public health model focuses on **prevention**. Rather than ending up having to treat disease, we want to prevent it from emerging in the first place. If alcoholism were a common problem, we would perhaps want increase taxation on alcoholic beverages. Moreover, we would perhaps recommend, or even enforce, public vaccination programs to avoid the spread of certain diseases. When we apply this ethical framework to public safety and criminal care, it would imply the realization that one evil does not correct another evil. Harshly punishing someone would require solid evidence.

Third, is the point that public health ethics rely on coercive **governmental intervention**, which means infringing upon individual rights to benefit the common good. Examples, which I have mentioned above, could perhaps be the enforcement of mandatory vaccinations, and quarantine.

Lastly, public health ethics are intrinsically **consequentialist**. This does not mean that they do not pay attention to the protection of individual rights. Nonetheless, one can imagine that conflicts between, for instance, the common good on one hand, and individual rights on the other, are inevitable:

> ..quarantine is only justified when (a) the benefits to society (protection from infectious disease) are greater than the burdens placed upon those in quarantine; (b) the burdens placed on those in quarantine cause the least harm possible; and (c) those placed In quarantine are provided with adequate care (including treatment). (Caruso, 2016, p. 41)

Surely, we can imagine cases where it is obvious should use quarantine, if they carried smallpox or malaria, for instance. Yet, would anyone at all think it is a good idea to quarantine someone with a common cold, or a treatable STD, in order to prevent infection to spread? That certainly sounds preposterous, because it would be an infringement of basic human rights, although with little to no benefit to justify this, and most likely, to a significant detriment to society because of the excessive economical cost such a practice would entail. However, one could perhaps imagine grey areas, where total isolation is not necessary, yet which calls for less invasive interventions.

So then, if this sort of model were to lay the groundwork for our criminal justice system, it would raise a need for major reform, according to (Caruso, 2016). It may for instance be easy to point to some simple, yet evidently suitable shifts in our justice system: A concrete example would perhaps be the end of the "war on drugs", given the evidence suggesting its harmful effects, as well as the remarkable lack of evidence suggesting any noteworthy desirable effects (Hughes & Stevens, 2010; Pedersen & Skardhamar, 2010). Hence, it might be reason to conclude that to incarcerate recreational drug users is not much different than quarantining someone with runny nose.

One might dispute, however, that this analogy is false, given that a drug user "knowingly breaks the law" whereas someone with a viral infection never saw it coming. This is of course an important point, as it would be frightening to live in a society where ones freedom could be taken away for the something as benign as a cold virus. Nonetheless, it is still quite an absurd argument, given that it just as well can justify laws that prescribe the most horrible punishments aimed at the most benign, but 'willed', behaviors, such as homosexuality, blasphemy and thought crimes, apostasy, adultery, or dancing. Many would perhaps include

"drinking wine" on that list, even though alcohol is shown to be more harmful than most, if not all illegal drugs (Nutt, King, & Phillips, 2010).

Hence, as a consequence of free will skepticism, it seems as if one is left with the burden of having to defend legal sanctions with empirical evidence, considered in light of basic human rights and the common good.

3.3 Replacing free will

So, when, how would we replace free will? Juth and Lorentzon (2010) argue that any scientific conception of personal autonomy, freedom, and responsibility should be understood as matters of degree, instead of a dichotomy. The authors continue to argue that, the term free will should be abandoned because of its historical connotations with retributivism, and suggest that a gradual conception of the term **autonomy** might be a good substitute.

On the other hand (see section 6.4), there has been offered good arguments in favor of keeping and redefining the free will-term in accordance with **Restorative Free Will**, again which is defined as: "*the capacity to effectively explore alternative paths in response to a combination of environmental contingencies and internal motives.." (Waller, 2015, p. 1).*

3.4 Clinical implications, and implications for further research

There are certainly potential limitations to consider, in regards to the arguments presented in this text. As I am familiar with the danger of motivated reasoning, and various cognitive biases, I know that mistakes can be made. For instance, even if my arguments proves to be sound, that would not matter much if I were arguing against a strawman. This problem, however, arguably sheds light on why the free will-discussion should not be treated as a purely philosophical issue, and shows why all scientific concepts should be analyzed and validated by scientists using empirical inquiry when necessary.

So then, if my arguments indeed are sound, then one question is whether the legal understanding of free will should be entirely abandoned, as suggested by Juth and Lorentzon (2010), or whether it should be revised, as suggested by Waller (2015). A problem with this latter proposal, however, is that Waller's revised conception of free will is quite different from the traditional, retributivist understanding currently found in most legal systems. This

might be a source of misunderstanding, which is why we might want to abandon the free will-concept altogether. As the philosopher Baruch Spinoza once wrote: “Many errors, of a truth, consist merely in the application of the wrong names to things.” (Ratner & de Spinoza, 1927, p. 190). Whether validated concepts like, locus of control (Rotter, 1966), self-efficacy (Bandura, 1977), or learned helplessness (Abramson et al., 1978), each can serve their purpose alone, or whether they should synthesize to a broader and gradual conception of autonomy, or free will, might be subject for further discussion.

All the same, I find good reason to expect that a rejection of the traditional free will axiom would enable forensic psychologists and psychiatrists to practice their profession much more as empirically grounded scientist, in alignment with their professional training. For instance, I suspect clinicians in the forensic setting would become less confined to the vagueness, or incoherence, of the legal insanity-concept, and more able to engage in the science of medicine and psychology, and more with scientifically validated concepts and reliable psychometrics, including violence risk assessments such as HCR-20 (Douglas, Hart, Webster, & Belfrage, 2013). Insofar the reliability of currently available assessments of this kind are poor, that would arguably not speak in favor of retributivism, but rather serve as a sign that we need further enquiry, in order to improve our assessments.

Moreover, I would expect that the abolition of retributivism would mean that clinicians no longer would have to break the ethics of the Hippocratic Oath: They would no longer be delegated the mandate to select individuals eligible or not for punishment used, in any degree, as harm for harm’s sake.

Perhaps psychiatrists and psychologists will have to lose their current authority in the courtroom, and instead serve as providers of more empirically validated information, as expert witnesses with no more authority than pathologists, social scientists, criminologists, and bioengineers. Without retributivism, it is in my view quite possible that legal insanity would become undermined as a concept entirely, as have been suggested by Norwegian forensic psychiatrist Rosenqvist (2012). This could perhaps require that we revise not only our criminal care system, but also our healthcare system, so that they could work more in cooperation, similar to the idea behind the Swedish criminal care system (Juth & Lorentzon, 2010). Rosenquist, however, seems to argue that forensic psychiatry rather should go toward

an opposing solution, by suggesting that clinicians might need to focus less on medical nosology for mental illness, and focus more on legal terminology (such as the tests for *legal* insanity). According to her, we may also need to elaborate further on potential criterions that ought to be present for a legal insanity-verdict. In my view, there might be good reason to dread such a development, as it would simply make the forensic psychiatric mandate more tedious and complicated, yet still as arbitrary, and no more coherent.

3.5 Conclusion

It is certainly difficult to envision all the clinical, legal and theoretical implications that eventually might have to follow the end of retributivism. It is however certainly beyond the scope of this text to find all answers to the many questions that seem to emerge from free will-skepticism. Nonetheless, it is important to note that the lack of an elaborate alternative theory of justice does not mean we should maintain the current system, and it does not refute my arguments. If it turns out that we have little to no empirical validation behind many of our laws, our current use of punishment, and the way we use compulsory hospitalizations, this should not make us fear utilitarianism, but rather embrace it. If there is a significant lack of evidence behind our current practices, it is arguably because we never had the obligation to justify them. Thus, it seems clear that a revision of our views on responsibility is needed largely because of-, and not despite the fact that we know so little about how to prevent crime, and how to isolate, treat and punish dangerous criminals.

In final, it seems clear that a revision or abandonment of the free will-concept, and the arrival of a system modeled akin to The Public Health-Quarantine Model (Caruso, 2016), would require that we use sound justification for all legal interventions, just as all medical interventions require today. It is no secret that medical interventions that are not based on empirical evidence are not regarded as medicine at all, but rather as pseudoscience. Perhaps, then, it is safe to say that the elimination of pseudoscience from our criminal care system, and from the current state of forensic psychiatry, would be most welcome. This would mean that we would have to cease many of our current practices, insofar they are unjustified, unproductive, and especially if they are shown to even be counterproductive. The fresh need for more empirically validated forensic assessments, and for new countermeasures against crime that are shown to be effective, on the other hand, would imaginably spur an insatiable

need for scientific enquiry within fields such as criminology, social science, economics, biology, medicine, and psychology.

References

Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of abnormal psychology, 87*(1), 49-74.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, text revision (DSM-IV-TR)*. (0890420629).

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied, 70*(9), 1-70.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review, 84*(2), 191-215.

Barash, D. P., & Lipton, J. E. (2011). *Payback: Why we retaliate, redirect aggression, and take revenge*: Oxford University Press.

Baumeister, R. F. (2002). Ego depletion and self-control failure: An energy model of the self's executive function. *Self and Identity, 1*(2), 129-136.

Becker, R. F. (2003). The evolution of insanity standards. *Journal of Police and criminal Psychology, 18*(2), 41-45.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological review, 74*(3), 183-200.

Berntzen, L. E., & Sandberg, S. (2014). The Collective Nature of Lone Wolf Terrorism: Anders Behring Breivik and the Anti-Islamic Social Movement. *Terrorism and Political Violence, 26*(5), 759-779.

Blass, T. (1999). The Milgram Paradigm After 35 Years: Some Things We Now Know About Obedience to Authority. *Journal of Applied Social Psychology, 29*(5), 955-978. doi:10.1111/j.1559-1816.1999.tb00134.x

Campbell, C. A. (1957). *On selfhood and godhood: the Gifford lectures delivered at the University of St.Andrews during sessions 1953-54 and 1954-55 revised and expanded*. London: Allen & Unwin.

Caruso, G. D. (2016). Free will skepticism and criminal behavior: A public health-quarantine model. *Gregg D. Caruso (2016)."Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model," Southwest Philosophy Review, 32*(1), 25-48.

Cashmore, A. R. (2010). The Lucretian swerve: the biological basis of human behavior and the criminal justice system. *Proc Natl Acad Sci U S A, 107*(10), 4499-4504. doi:10.1073/pnas.0915161107

Christie, N. (2007). *Limits to pain: The role of punishment in penal policy*: Wipf and Stock Publishers.

Darwin, C., Barrett, P. H., & British, M. (1987). Charles Darwin's notebooks, 1836-1844 : geology, transmutation of species, metaphysical enquiries. Cambridge: Cambridge University Press.

Delgado, J. M. R. (1969). *Physical control of the mind: Toward a psychocivilized society* (Vol. 41): World Bank Publications.

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied psychology, 86*(3), 499-512.

Dennett, D. C. (1981). *Brainstorms: Philosophical essays on mind and psychology*: MIT press.

Dennett, D. C. (1984). *Elbow room: the varieties of free will worth wanting*. Oxford: Clarendon Press.

Dennett, D. C. (2004). *Freedom evolves*. London: Penguin Books.

Descartes, R., & Cottingham, J. (2013). *René Descartes: Meditations on first philosophy: With selections from the objections and replies*: Cambridge University Press.

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20v3: Assessing risk for violence: User guide*: Mental Health, Law, and Policy Institute, Simon Fraser University.

Ebert, J. P., & Wegner, D. M. (2011). Mistaking randomness for free will. *Conscious Cogn, 20*(3), 965-971. doi:10.1016/j.concog.2010.12.012

Faden, R., & Shebaya, S. (2015). Public Health Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015 ed.).

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology, 58*(2), 203.

Fischer, J. M., Kane, R., Pereboom, D., & Vargas, M. (2009). *Four views on free will*: John Wiley & Sons.

Fleishman, E. A., & Harris, E. F. (1962). Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology, 15*(1), 43-56.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 68*(1), 5-20.

Gesch, C. B., Hammond, S. M., Hampson, S. E., Eves, A., & Crowder, M. J. (2002). Influence of supplementary vitamins, minerals and essential fatty acids on the antisocial behaviour of young adult prisoners Randomised, placebo-controlled trial. *The British Journal of Psychiatry, 181*(1), 22-28.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological bulletin, 117*(1), 21.

Glueck, S. (1962). *Law and psychiatry: cold war or entente cordiale?* : Johns Hopkins Press.

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philos Trans R Soc Lond B Biol Sci, 359*(1451), 1775-1785.

Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. International Journal of Criminology and Penology, 1, 69-97.

Hausenblas, H. A., Carron, A. V., & Mack, D. E. (1997). Application of the Theories of Reasoned Action. Journal Of Sport & Exercise Psychology, *19*, 36-51.

Hughes, C. E., & Stevens, A. (2010). What can we learn from the Portuguese decriminalization of illicit drugs? *British Journal of Criminology*, 50, 999-1022.

Isen, A. M., & Levin, P. F. (1972). Effect of feeling good on helping: cookies and kindness. *Journal of personality and social psychology, 21*(3), 384.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science, 310*(5745), 116-119.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*(1), 1-24.

Juth, N., & Lorentzon, F. (2010). The concept of free will and forensic psychiatry. *Int J Law Psychiatry, 33*(1), 1-6. doi:10.1016/j.ijlp.2009.10.008

Kant, I. (1996). *Kant: The metaphysics of morals* (M. J. Gregor, Trans. M. J. Gregor Ed.): Cambridge University Press.

Kant, I. (2006). *Kant: anthropology from a pragmatic point of view*: Cambridge University Press.

Kavanau, J. L. (1967). Behavior of captive white-footed mice. *Science, 155*(3770), 1623-1639.

Kim, J. (2000). *Mind in a physical world: An essay on the mind-body problem and mental causation*: MIT press.

Latorre, J. I. (2013). Quantum Will: Determinism meets Quantum Mechanics. *Euresis journal, 5*, 1-16.

Leotti, L. A., & Delgado, M. R. (2011). The Inherent Reward of Choice. *Psychological Science, 22*(10), 1310-1318.

Lerner, M. J. (1980). *The belief in a just world*: Springer.

Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies, 6*(8-9), 47-57.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain, 106*(3), 623-642.

Melle, I. (2013). The Breivik case and what psychiatrists can learn from it. *World Psychiatry, 12*(1), 16-21.

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371.

Milgram, S., & Van den Haag, E. (1978). Obedience to authority: Ziff-Davis Publishing Company.

Monroe, A. E., Dillon, K. D., & Malle, B. F. (2014). Bringing free will down to Earth: people's psychological concept of free will and its role in moral judgment. *Conscious Cogn, 27*, 100-108. doi:10.1016/j.concog.2014.04.011

Moore, M. S. (2009). *Causation and responsibility: An essay in law, morals, and metaphysics*: Oxford University Press.

Moore, M. S. (2012). Responsible choices, desert-based legal institutions, and the challenges of contemporary neuroscience. *Social Philosophy and Policy, 29*(01), 233-279.

Moore, M. S. (2014). The Quest for a Responsible Responsibility Test: Norwegian Insanity Law After Breivik. *Criminal Law and Philosophy*. doi:10.1007/s11572-014-9305-6

Nahmias, E. (2011). Why 'willusionism'leads to 'bad results': Comments on Baumeister, Crescioni, and Alquist. *Neuroethics, 4*(1), 17-24.

Nahmias, E., Morris, S. G., Nadelhoffer, T., & Turner, J. (2006). Is Incompatibilism Intuitive? *Philosophy and Phenomenological Research, 73*(1), 28-53. doi:10.1111/j.1933-1592.2006.tb00603.x

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review, 84*(3), 231.

NOU 2014:10 (2014). *Skyldevne, sakkyndighet og samfunnsvern*. (NOU 2014:10,). Oslo: Departementenes sikkerhets- og serviceorganisasjon, Informasjonsforvaltning.

Nutt, D. J., King, L. A., & Phillips, L. D. (2010). Drug harms in the UK: a multicriteria decision analysis. *The Lancet, 376*(9752), 1558-1565. doi:10.1016/S0140-6736(10)61462-6

O'Connor, S., & Vandenberg, B. (2005). Psychosis or faith? Clinicians' assessment of religious beliefs. *Journal of Consulting and Clinical Psychology, 73*(4), 610.

Pedersen, W., & Skardhamar, T. (2010). Cannabis and crime: findings from a longitudinal study. *Addiction, 105*(1), 109-118.

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

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*: Oxford University Press.

Pierre, J. M. (2014). The neuroscience of free will: implications for psychiatry. *Psychol Med, 44*(12), 2465-2474. doi:10.1017/S0033291713002985

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American journal of health promotion, 12*(1), 38-48.

Pärnamets, P., Johansson, P., Balkenius, C., Hall, L., Spivey, M. J., & Richardson, D. C. (2013). *Changing minds by tracking eyes: Dynamical systems, gaze and moral decisions.* Paper presented at the CogSci 2013.

Ratner, J., & de Spinoza, B. (1927). *The Philosophy of Spinoza: Selected from His Chief Works*: Modern Library.

Rosenqvist, R. (2012). Utilregnelighetsregelen—moden for revisjon? *Tidsskrift for den Norske Laegeforening, 132*(7), 843.

Ross, D., & Spurrett, D. (2004). What to say to a skeptical metaphysician: A defense manual for cognitive and behavioral scientists. *Behavioral and brain sciences, 27*(05), 603-627.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied, 80*(1), 1.

Ryle, G. (1949). *The Concept of mind*. New York: Barnes.

Sanvik, S., & Nilssen, I. D. (2013). Derfor kan kriminelle gå fri for trusler, vold, overgrep og drap. Retrieved from https://www.nrk.no/norge/voldsmenn-slipper-dom-1.10864943

Sartre, J.-P. (1989). *Being and nothingness: an essay on phenomenological ontology*. London: Routledge.

Schopenhauer, A., & Kolenda, K. (2012). *Essay on the Freedom of the Will*: Dover Publications.

Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., . . . Vohs, K. D. (2014). Free will and punishment: a mechanistic view of human nature reduces retribution. *Psychol Sci, 25*(8), 1563-1570. doi:10.1177/0956797614534693

Silk, J. B. (2016). Evolution: Taxonomies of cognition. *Nature, 532*(7598), 176-176. doi:10.1038/532176a

Skeie, C., & Rasmussen, K. (2015). Assessment of causal associations between illness and criminal acts in those who are acquitted by reason of insanity. *Tidsskrift for den Norske laegeforening: tidsskrift for praktisk medicin, ny raekke, 135*(4), 327.

Smilansky, S. (2000). *Free will and illusion*: Oxford University Press.

Smilansky, S. (2011). Hard determinism and punishment: A practical reductio. *Law and Philosophy, 30*(3), 353-367.

Steen, B. H. (2012, 22.08.2012). Slik skal Breivik bo. Retrieved from http://www.tv2.no/a/3857512

Syse, A. (2006). Strafferettslig (u) tilregnelighet-juridiske, moralske og faglige dilemmaer. *Tidsskrift for strafferett*(3), 141-175.

Syse, A. (2012). Punishment, treatment and fair retribution. *Tidsskr Nor Laegeforen, 132*(7), 841-843. doi:10.4045/tidsskr.12.0344

Syse, A. (2014). Breivik – The Norwegian Terrorist Case. *Behav Sci Law, 32*(3), 389-407. doi:10.1002/bsl.2121

Taylor, P. J. (1985). Motives for offending among violent and psychotic men. *The British Journal of Psychiatry, 147*(5), 491-498.

Thorvik, A. (2000). Tema-Rettspsykiatri-Frihet som medisinsk problem-rettspsykiatri i filosofisk lys. *Tidsskrift for Den norske lægeforening, 120*(18), 2154-2158.

Thorvik, A. (2008). Rett og urett-Strafferettslig utilregnelig--vitenskap eller skjonn? *Tidsskrift for den Norske legeforening, 128*(16), 1864.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: encouraging a belief in determinism increases cheating. *Psychol Sci, 19*(1), 49-54. doi:10.1111/j.1467-9280.2008.02045.x

Waller, B. N. (2011). *Against moral responsibility*: MIT Press.

Waller, B. N. (2014). *The Stubborn System of Moral Responsibility*: MIT Press.

Waller, B. N. (2015). *Restorative free will: Back to the biological base*: Lexington Books.

Weiner, B. (1993). On sin versus sickness: A theory of perceived responsibility and social motivation. *American Psychologist, 48*(9), 957.

Wilson, T. D., & Dunn, E. W. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Psychology, 55*.

World Health Organization. (1999). *ICD-10: Psykiske lidelser og atferdsforstyrrelser: Kliniske beskrivelser og diagnostiske retningslinjer*.