

# **Responsibility, Blame and the Psychopath**

A Dissertation Presented for the  
Doctor of Philosophy  
Degree  
The University of Tennessee, Knoxville

Matthew William Ruble  
May 2015

Copyright © 2014 by Matthew William Ruble  
All rights reserved.

To My Family and all my playmates  
Especially Ella

*A few hours of mountain climbing make a scoundrel  
and a saint two fairly equal creatures.  
Fatigue is the shortest path to equality and fraternity –  
and sleep finally adds to them, liberty.*

-Nietzsche

## **Acknowledgments**

I am grateful to the Philosophy Department of the University of Tennessee for the opportunity provided to me through a graduate teaching assistantship. I have received invaluable professional and personal support from many members of the department during my studies. I thank Dr. David Reidy, Dr. David Palmer, and Dr. Lois Presser for serving on my dissertation committee. I would also like to thank the co-directors of this dissertation, Dr. Adam Cureton and Professor Emeritus Dr. John Hardwig. I consider myself doubly fortunate to have worked with both Adam and John on this dissertation. Both are exemplars of intellectual generosity. I would also like to thank Susan Williams and Cheryl Smith for their invaluable support.

I am also deeply appreciative of the financial support I received from The Prados Fund 2014 summer value theory dissertation fellowship. All the arguments and errors expressed in this dissertation are my own and are not reflective of the views of The Prados Fund.

To my family and friends for all your love and support, thank you for sticking with me and for encouraging me to stick with it. You are my motivation.

## Abstract

This dissertation examines the moral responsibility of psychopaths. I begin with an analysis of the concept of psychopathy by situating it within the context of a central debate in the philosophy of psychiatry over the conceptual nature of mental illness to demonstrate that psychopathy is an inherently value-laden concept. I argue against the disease-model of psychopathy and against their automatic exemption from moral responsibility as argued for by many moral philosophers. Psychopaths possess sufficient agency such that exempting them from moral responsibility is problematic both epistemically and morally. Yet psychopaths frequently offer reasons for their behavior that reveal their distance from full moral agency. So how are we to respond to such middle ground moral agents when they do terrible things to other people? The discussion then turns to the normative question of how we should respond to wrongdoing psychopaths. This analysis begins with the framework for general responses to moral wrongdoing as provided by moral philosopher P.F. Strawson. The enduring distinction between the reactive attitudes and the objective view is challenged as overly coarse and potentially morally disrespectful to the mentally ill, including specifically, psychopaths. I conclude that the question of whether or not psychopaths are 'in fact' morally responsible remains open and thus, forces us to take up the question of how we should understand our responses to psychopathic wrongdoing, including whether or not we can or should hold psychopaths responsible. The case of psychopaths reveals to us alternative ways of understanding how we 'hold responsible,' beyond resentment and blame for example, that do not lose moral content simply by straying from the paradigmatic reactive attitudes. If we forgo the blame paradigm of moral responsibility we find morally sensible ways of responding to wrongdoing psychopaths.

## Table of Contents

<b>Introduction .....</b>	<b>1</b>
<b>Chapter 1: Conceptual models of psychopathy .....</b>	<b>5</b>
1.0 Introduction .....	5
1.1 A central debate in the philosophy of psychiatry.....	10
1.2 Two accounts of psychopathy: value-free and value-laden .....	13
1.3 The empirical argument .....	18
1.4 No escaping values: criticism of the empirical argument.....	20
1.5 Antisocial Personality Disorder: value-laden diagnostic criteria .....	27
1.6 Cases .....	36
1.7 Discussion of Cases .....	39
1.8 Conclusion .....	54
<b>Chapter 2: Strawson’s framework .....</b>	<b>59</b>
2.1 Introduction.....	59
2.2 Strawson’s framework .....	60
2.2.1 Structure of the interpersonal reactive attitudes.....	62
2.2.2 Commentary on the reactive attitudes.....	64
2.3 On the reactive attitude/objective view distinction .....	67
2.4 Concerns with Strawson’s account.....	75
2.4.1 Agency recognition and the tension between the reactive attitudes and the objective view .....	76
2.4.2 On the general propriety of reactive attitudes .....	78
2.4.3 On the general propriety of the objective view .....	90
2.5 Conclusion .....	96
<b>Chapter 3: Levy’s view .....</b>	<b>98</b>
3.1 Introduction: Levy’s general view and its appeal.....	98
3.2 Causal responsibility and morally responsibility .....	100
3.3 The epistemic condition for moral responsibility: understanding, reasons, and ‘pathological’ reasons .....	102
3.4.1 Levy’s inversion.....	111
3.4.2 Levy’s response to wrongdoing psychopaths.....	114
3.5 Conclusion .....	120
<b>Chapter 4: Beyond the common view .....</b>	<b>121</b>
4.1 Introduction.....	121
4.2 Two types of <i>reactans</i> : holding and expressing the reactive attitudes.....	124
4.3 Two types of <i>reactanda</i> : agents and behavior.....	128
4.4.1 The common view’s narrow <i>reactans-reactandum</i> pairing .....	134
4.4.2 Is gratitude prohibited too? .....	138
4.5 Objections .....	140
4.6 Conclusion .....	143
<b>Conclusion.....</b>	<b>144</b>
<b>Bibliography.....</b>	<b>152</b>
<b>Vita.....</b>	<b>158</b>

## **Introduction**

This dissertation takes up the question of the moral responsibility of psychopaths, including the question of how we should respond to wrongdoing psychopaths. Although several prominent moral philosophers are engaged herein, this dissertation is written from the perspective of the philosophy of psychiatry. The debate over psychopathic responsibility is thus enriched by a greater appreciation of the philosophical challenges presented by the concepts of mental illness, including the concept of psychopathy, an alleged illness of moral agency. Moral philosophers frequently appear to presuppose that psychopathy, along with various other mental illnesses, are sufficiently like physical diseases such that those inflicted lack the requisite control over their actions typically believed to ground moral responsibility, (Levy, 2007, 2010; Strawson, 1993; Watson, 1993; Wolf, 1987; Elliot, 1996, Fisher & Ravizza, 1998; Fischer and Tognazzini, 2013). The views on the responsibility of psychopaths held by moral philosopher, P.F. Strawson, and neuroethicist, Neil Levy are closely and directly engaged in this dissertation.

Accordingly, this dissertation attempts to navigate two equally difficult concepts that we often find mutually offsetting: mental illness and moral responsibility. This task is complicated by the tendency of moral philosophers to cite some specific mental illness (typically schizophrenia, addiction or psychopathy) as intuitively obvious and paradigmatic cases of non-morally responsible agency. As a former mental health professional and as a philosopher specifically interested in conceptual and ethical issues that surround mental illness, I find such analyses and

conclusions problematic. We need to confront the fact that many mental illnesses may not represent a total, or even partial, exemption from moral responsibility. In our confrontation, however, we need to be clear about exactly what practices of 'holding' mentally ill persons responsible we are warranting, and why. We do not want to insist that the mentally ill are morally responsible simply so that we are free to harshly judge, cathartically blame and resent the ill wrongdoer. Harsh judgment and resentment may be practices that we need to forego more generally toward the healthy and ill alike. We want to be very cautious in dismissing the moral status of the mentally ill so that we remain vigilant against the many forms of disrespect and dehumanization that comes with objectifying the mentally ill. By seeking to protect the mentally ill from moral blame we may well bring more harm than we sought to avoid.

In this dissertation I argue in favor of the view that psychopaths possess sufficient agency to be considered morally responsible for their wrongdoing. At the very least, I favor the view that the burden of proof that the empirical basis alleged to support the claim that psychopaths are not responsible has thus far failed to make its case. However, clearly establishing the case that psychopaths are 'in fact' responsible must not rely on the assumption that holding psychopaths responsible simply licenses a commitment to the affective expression of resentment. There is a more robust sense of 'holding responsible' that is not limited to affectively charged blame and resentment. This dissertation thus offers ways of holding psychopaths responsible that are beyond the dispositional stereotype of blame and resentment. We are all too aware of the unedifying spectacle of moralizing blame and how

sustained resentment can sour a soul. Even if some psychopaths came to be psychopaths through no fault of their own, say as a result of severe childhood abuse or a broken region of their brain, (a conjecture that is far from established), then we should not be insensitive to this in gauging our responses to their ongoing wrongdoing. Taking up such a morally insensitive disposition to the psychopath would put us in the ironic position of exhibiting moral insensitivity, a trait that alleges to define psychopathy itself. There is, however, a good case to be made that we can hold the psychopath responsible for his wrongdoing and that we can do so from both an informed and morally sensitive approach. I aim to defend such a view in this dissertation.

This task is complicated by a set of unclear concepts all of which will be scrutinized in this dissertation, including: the nature of the *moral* illness of psychopathy, and the twin questions of whether or not a person 'is' responsible and whether or not we should 'hold' her responsible. Although I devote some attention to the question whether or not the psychopath 'is' responsible by challenging the disease status of psychopathy, I focus more closely on the analyzing the question of 'holding' the psychopath responsible, and which of the meanings of 'hold responsible' are appropriate or justified with respect to our responses to psychopathic wrongdoing. The common view among moral philosophers is that we must first answer the question whether or not the psychopath is 'in fact' responsible and only after which can we take up the normative question of how we should respond to his wrongdoing, including variations of what it means to 'hold' persons responsible. Thus, my contribution can be thought of as attending to the 'hold

responsible' question that has been largely neglected in much of the discussion regarding the moral responsibility of psychopaths.

The common view also holds that psychopaths are completely morally blind (alternatively, psychopaths categorically lack moral understanding) and that they are incorrigible in that there is no possibility or hope for moral improvement, however miniscule the gains. This pessimistic view is underwritten by the disease-model conception of psychopathy that seeks to explain this phenomenon in medical and empirical terms. Here, the perspective from the philosophy of psychiatry stands to offer considerable insight into the situation by challenging the conceptual assumptions of psychopathy that moral philosophers assume as a starting point in their discussion of the moral responsibility of psychopaths. If we attend more closely to these conceptual assumptions of psychopathy we find that the disease model is far from established and without this assumption, the complexion of psychopathic responsibility shifts significantly. Hence, when engaging the question of psychopathic responsibility, we are dealing with two normative concepts.

After summarizing the common view on the moral responsibility of psychopaths held by the moral philosopher P.F. Strawson and neuroethicist Neil Levy, I expose an optional understanding of 'holding' psychopaths responsible that avoids a key concern held by both philosophers while exposing the moral problems of the response called for by the common view, that we should isolate and indefinitely detain the wrongdoing psychopath. Such disengagement does not represent a morally enlightened response.

# **Chapter 1: Conceptual models of psychopathy**

## **1.0 Introduction**

This chapter provides an overview of many themes related to psychopathic responsibility that will be explored and clarified in greater depth throughout the dissertation. Central among the themes of this chapter is the conceptual model that views psychopathy as moral disease. Philosophers customarily skip this analysis en route to the conclusion that the mental disorder renders psychopaths not morally responsible. This is an omission that, when filled by closer conceptual examination, may very well alter such a conclusion. We will find that contrary to the common view among philosophers, psychopaths (or those persons so labeled) are not necessarily completely morally blind nor do they suffer from an incurable disease.

Just what is a psychopath? Are psychopaths a natural kind, perhaps akin to what we sometimes describe as a force of nature such as tornados and earthquakes? But psychopaths are human beings and we typically refer to persons as 'forces of nature' only in a metaphorical sense. Perhaps psychopathy is a moral disease, (odd as this concept sounds), that is itself a force of nature acting on, and through, unsuspecting agents. But this is to imbue psychopathy with an independent ontology distinct from the person over which it has a grip. Perhaps what philosophers call a 'psychopath' is merely a counterfactual, or hypothetical, critter used as a tool to better understand the nature of moral responsibility. But counterfactuals and stipulating features (like complete and terminal moral blindness) ignore at least two things: that those diagnosed with psychopathy (or Antisocial Personality Disorder to use the clinical title) are real people and it ignores

key evidence, as little as there is, that we have about actual people diagnosed with this disorder. In this case philosophers should speak of the moral responsibility of 'counterfactual critters' rather than 'psychopaths.' And since the diagnostic clinical term *that people are diagnosed with* is Antisocial Personality Disorder, not 'psychopathy,' it may very well be the case that a 'psychopath' is nothing more than a name philosophers have given to this counterfactual critter. But we want to know how we should respond, morally speaking, to wrongdoing committed by real people who have been diagnosed with this disorder. We want to know if we should at least attempt to hold them morally responsible.

My own use of the term 'psychopath' refers to an actual class of people diagnosed with Antisocial Personality Disorder, rather than a stipulated philosophical concept that presumes moral blindness and moral incorrigibility. We want to know as a practical matter whether or not these real people are morally responsible and how, if at all, we should go about holding them responsible. Looking then to psychologists and psychiatrists for a better understanding of psychopathy (Antisocial Personality Disorder), we find it initially described as a steady pathological state of one's personality. In this sense, the term 'psychopath' alleges to describe core features of an agent's *personhood*. But 'psychopathy' can only be diagnosed by way of behavioral criteria, which we also refer to as 'symptoms.' Perhaps then, psychopathy is a conceptual device and a term of convenience used to distinguish persons according to behavioral patterns (in this case patterns of criminally violent behavior). But this does little to aid our understanding of the psychopath's status as a moral agent and says less about his

ability to ever improve on the pathological state of his morality. And so we find the psychologists understanding of psychopathy inadequate with respect to conceptual clarity and unequipped to answer questions about the moral responsibility of these persons.

Embedded in these understandings of psychopathy are assumptions, conceptual and *causal* assumptions, about the nature of psychopathy that have direct bearing on the question of the responsibility of psychopaths. Since we do not blame forces of nature like hurricanes and tornados for the destruction they cause, (though we may very well find human responses to such phenomena that include blame or praise), whether or not we blame psychopaths for the destruction they cause will depend on our conceptual understanding of psychopathy. If psychopaths are indeed a similar force of nature, then we should similarly not blame them for the destruction they cause. But some might find this analogy rather strained. If psychopaths are, however, unlike more typical forces of nature, then what are they? As I hope to show in this chapter, when we ask this question, are we asking a conceptual question in which both empirical facts and moral values are inexorably involved.

The main aim of this chapter is to demonstrate that far from being value-free empirical concept (like physical disease) psychopathy is morally value-laden concept. On my understanding, psychopathy is informed by both empirical and moral concepts and is thus, conceptually thick. There is considerable conceptual ground to cover to show this conceptual complexity and to additionally show why this matters for understanding the responsibility of psychopaths. Since we

commonly appeal to historical concerns in assessing the responsibility of a given agent (or action), the etiology of psychopathy is pertinent to such assessment. There is also the matter of addressing significant empirical evidence in support of the view that psychopaths lack the requisite causal control over their actions required for moral responsibility. (In this dissertation, empirical evidence will be handled in two stages. The first stage comes in this chapter's conceptual analysis. The second comes later in chapter 3 which engages more detailed and recent evidence cited by Levy to support the conclusion that psychopaths are not responsible.) The general argument that psychopaths are not responsible assumes psychopathy to be a disease with a causal etiology. The common view argument can be summarized as follows. If psychopaths do not cause the onset of their psychopathy (*qua* moral disease), and psychopathy causes violent behavior, then psychopaths are not the cause of their actions (including violent actions) in a way that one must be the cause of their actions in order to be morally responsible. Although this argument comes with some variety, many philosophers share the general conclusion and take this view quite seriously. But they do so without paying much attention to the *concept* of psychopathy and either assume or stipulate<sup>1</sup> that psychopaths suffer completely and incorrigibly from 'moral blindness'. In the minds of many philosophers, psychopaths suffer from an incurable disease that adversely affects the morality of those so diagnosed. But once we show that the concept of psychopathy has an alternative value-laden understanding, that the concept itself

---

<sup>1</sup> Strawson (1993), Watson (1993), Levy (2007), Talbert (2008), Fischer and Ravizza (1998), stipulate that psychopaths are categorically morally blind and diseased.

appears to be morally value-laden, then the view of the psychopath as a kind of diseased agent (in a strictly empirical sense) that is assumed by so many philosophers, is unsatisfactory. And we do not want to reach a verdict on the responsibility of the psychopath based on an unsatisfactory understanding of psychopathy.

We must also keep in mind that even if we reject a value-free disease model of psychopathy it does not straightforwardly follow that psychopaths are responsible enough to justify blaming them (either by our internal convictions or expressions of those convictions in our reactive attitudes). Perhaps psychopaths really are responsible but we ought not blame them, or even feel negative reactive attitudes toward them. But if they are in some degree morally responsible, then perhaps we are justified in a measured degree of blame. There is a vast middle ground between fully responsible moral agents (adults) and fully not-responsible 'agents' (children), and perhaps the manner in which we hold the middle-ground agents (adolescents) responsible should be suitably tailored. We must also recognize that the degree of extant responsibility of a given agent at a given time need not simply be symmetrically matched with a likewise fullness of moral response. Well-tailored reactions to moral wrongs by such middle ground agents may call for demand more than an agent is currently capable of meeting. Sometimes we need be more considerate of the diachronic nature of agents when expressing our reactive attitudes. Doing so may also be a measure of respect for middle ground agents.

What does our conceptual understanding and model of psychopathy have to do with the responsibility of psychopaths? First, it is empirically the case that the disease model of *any* mental illness, psychopathy included, impacts our understanding of the responsibility of the mentally ill. This reflects a central conceptual debate in the philosophy of psychiatry that will be briefly discussed in the following section. Experimental philosophy has proven fruitful in demonstrating the tendency of this belief in the general population.<sup>2</sup> Many philosophers who comment on the moral responsibility of the mentally ill reflect this tendency. However, when conceived solely through the disease model, (or, when viewed as a disease) mental illness is uncritically taken as an intuitively obvious exemption from responsibility. But this may not be true and we need to confront the fact that mental illness does not provide a categorical exemption from responsibility. Perhaps there are subtle and unique ways in which specific mental illnesses affect moral responsibility but without removing it entirely. To investigate this, below I offer some cases involving different mental illnesses in order to probe how our intuitions position the psychopath, along with other mental illnesses, along the continuum of responsibility.

### **1.1 A central debate in the philosophy of psychiatry**

The concept of mental illness remains essentially contested and the philosophical debate has yielded two distinct general accounts. The descriptive

---

<sup>2</sup> See Knobe (2008)

account (disease-model) holds that mental illness, like physical illness, can be explained in objective, value-free terms of, say, a failure of biological functioning (Kendall, Boorse, Wakefield).<sup>3</sup> Normative accounts range from the stronger, anti-psychiatrists (Szasz, Foucault, Laing), to weaker non-descriptivist accounts (Fulford, Thornton, Nordenfelt, Edwards).<sup>4</sup> To varying degrees all normative accounts acknowledge the inescapable role of values (social, legal and moral norms) involved in classifying mental conditions and behavioral problems as discretely diagnosable mental illnesses.<sup>5</sup>

The normative camp in the mental illness debate resists the extension of a strictly biological account of physical illness to mental illnesses. The rejection of the analogy amounts to this: while the focus on specific organ *function* provides a useful and relatively simple account of physical illness as objectively factual and independent of values, it does not provide a similarly simple value-free account of mental illness because the practice of such classifications are based on inherently value-laden assessments of behaviors rather than, for example, reference to a lesion of an organ.<sup>6</sup> Diagnostic criteria for virtually all psychological disorders listed in the *DSM IV TR* confirm this worry.<sup>7</sup> No fMRI or genetic testing is required for someone

---

<sup>3</sup> Kendall, R.E. (1975); Boorse, C. (1997); Wakefield, J.C. (1992, 2009)

<sup>4</sup> Szasz, T. (1974, 1998); Fulford, KWM (1989); Nordenfelt, L. (1987); Edwards, C. (2009); Thornton (2007)

<sup>5</sup> For a relevant example of an evaluative criterion, see A.1 for Antisocial Personality Disorder: “failure to conform to social norms with respect to lawful behaviors.” Thus, no quick dismissal of the role of social values in ASPD is available.

<sup>6</sup> See Szasz (1978, 1998, 2006) for the strongest version of this argument

<sup>7</sup> DSM IV TR (2000): The diagnostic and statistical manual of mental disorders, fourth edition, text revision. The DSM, among other influences it has, sets the terms and standards for how various mental disorders are defined and classified. Since the

to be diagnosed with a psychological disorder.<sup>8</sup> Rather, behavioral criteria continue to solely serve this purpose. Because the practice of diagnosing mental illnesses concerns persons and their behavior rather than organs and their function, the actual reference of the symptoms is to (socially dysfunctional) behavior rather than to discrete biological dysfunctions. Furthermore, those espousing the normative account worry that the disease model runs the risk of ‘medicalizing the moral’ and redefining the boundaries of our moral landscape on conceptually mistaken grounds. Much of this debate has centered on the psychotic disorders (schizophrenia and the problem of ‘delusion’), where the intuitions of many tend to lean more strongly toward the disease model. And if the psychotic disorders (e.g., schizophrenia) are conceptually controversial (with regard to the accuracy of the disease model), then the personality disorders (especially one that alleges to describe a diseased moral character) appear anything but a value-free disease. The very idea of a value-free concept of a diseased personality is caught awkwardly straddled between the natural-normative divide.<sup>9</sup>

---

most recent philosophical writing pertaining to psychopaths occurred prior to DSM V, I use DSM IV criteria where the discussion is appropriate.

<sup>8</sup> fMRI: functional magnetic resonance imaging

<sup>9</sup> That social norms have played a role specifically with the Axis II Personality Disorders, one need look no further than the American Psychological Association’s lamentable history with what was once an Axis II disorder, ‘Homosexual Personality Disorder,’ present until the *DSM III R*, when it was rightly excised as a result of relentless social pressure in the mid 1980s. Furthermore, there is an on-going debate about whether or not to include *any* Axis II personality disorder in the forthcoming *DSM V*. I for one am not confident that personality disorders will be removed. And so, there is a meta-classification question lurking here that I cannot treat properly in this paper, to wit. Why even classify the behavioral criteria (symptoms) of ASPD as evidence of a disorder at all? Why medicalize such morally-laden behaviors? This touches a much deeper history of the theoretical

## 1.2 Two accounts of psychopathy: value-free and value-laden

In the introduction to this chapter I noted that philosophers assume that psychopaths are completely morally blind and that they will inevitably remain so impaired. I also suggested that this view arises, in part, from an underlying assumption that psychopathy is a disease and that the disease is the cause of his violent behavior in such a way that renders him not morally responsible. Philosophers share this general view, though with some variation of how the disease of psychopathy divorces those suffering from it, psychopaths, from responsibility.<sup>10</sup> Since the disease view of psychopathy appears to prop up the view that psychopaths are not morally responsible it warrants special analysis. As I shall argue, since psychopathy (viz. Antisocial Personality Disorder) is maintained with explicitly moral terms it is, far from being a value-free medical and empirical concept, is rather a value-laden concept in which moral norms cannot be reduced out of the conceptual picture. As such, psychopathy is uniquely situated within a larger and more comprehensive debate over the nature and concept of mental illness and the status of values contained therein. Below, I briefly summarize that debate and explain what bearing it has on questions about responsibility and mental illness.

---

development of all psychopathologies, but for a nice introduction to a set of related concerns, see Elliot (2008).

<sup>10</sup> Wolf's (1993) focus is on the concept of 'deep self' whereas Fischer and Ravizza (1998) focus on the concept of 'control.' Both accounts attribute the lack of these desiderata for responsibility to the disease of psychopathy to explain why the person with psychopathy is not responsible.

There are roughly two competing theories of mental illness, the value-free model (also referred to as the 'disease model' or the 'medical model') that purports to define mental illness strictly in value-free empirical terms (only after which, and separately, come ethical concerns) and the value-laden model that contends that the concept of mental illness itself (including various diagnostic criteria) is permeated with social, legal and ethical values.

I am not as concerned here with *how* values are smuggled into the concept of mental illness, specifically with respect to Antisocial Personality Disorder, so much as I am interested in showing *that* values permeate the diagnostic criteria, and thus, the very concept of psychopathy. (Later, in chapter 3 we will see how Levy unwittingly smuggles values into his assessment of psychopaths' responses to variations of the Trolley Problem.) In fact, the presences of values (moral values) in the diagnostic criteria for Antisocial Personality Disorder are far more centrally infused than that of other mental illnesses. I do, however, want to say just a few things about 'how' values come to be found in the concept of mental illness and many of its diagnostic criteria.

Philosophers of psychiatry often speak of the 'smuggling' of values into both the core concept of mental illness and many of its diagnostic criteria. This, however, may be subtly misleading by suggesting that those tasked with (re)defining various experiences in living as psychological disorders, (APA committees; for instance, the (re)defining of habitual violent behavior as a psychological disorder; or defining various sexual experiences as pathological or not – we should remember that just two decades ago homosexuality was a clinical psychological disorder; this error has

since been rectified), are consciously intentional in installing values in the criteria. This seems both unlikely and uncharitable. Rather, there may well be a conscious effort to purge the concept of mental illness and many of its criteria of values that has routinely failed and the fault lies in failing to recognize that the values purge is not possible. The unfortunate result then is that moral values become sublimated into the diagnostic criteria leaving the false impression of a successful values purge. So, rather than worrying about values being smuggled in, we should worry about how values, specifically what we call moral values, remain sublimated. A better option is to be open and transparent about the presence of values in the concept of mental illness so that the debate we have is clearly a normative debate (say, about our social policies for those with mental illnesses) that at its very core, is likewise normative in content.

For our purposes, psychopathy, perhaps more than any other discrete mental illness, shows this. In addition to showing this, I will also argue that rather than allowing these moral values defining psychopathy to remain sublimated, thus perpetuating the pretense of psychopathy as a disease, we should be explicit in our discussion about the values used to define the very idea that psychopathy is a disorder. Psychopathy is thus a deeply confused concept that leads to the erroneous notion that these are persons wholly incapable of moral responsibility rather than say, these are persons unusually efficient in failing to meet even basic moral demands not to assault and kill others, for instance. Levy (2007, 2010), like psychiatrist Hervey Cleckley (1955), (the conceptual progenitor of 'psychopathy'), is motivated to pathologize habitually violent behavior because they both see

psychopaths as victims of a disease and that it follows from this that we should not see them as moral agents worthy of responses that have moral content.

To see why this is a problem, let's compare psychopathy again to homosexuality. When in the mid 1980's, the APA dropped homosexuality personality disorder from the DSM, thus removing homosexuality from the list of mental illnesses, and thus, from being viewed (at least by clinicians, and perhaps to much of the public) as a disease. It may be that homosexuality was first included in the DSM from a compassionate motive to dissuade moral judgment (including reactive attitudes). Regardless of how homosexuality came to be viewed as a disease in the first place, what matters is the social and moral progress that occurred when it was removed from the list of mental illnesses. Of course this means that homosexuals once again became targets of moral condemnation (as if this ever ceased in the religious milieu, even during the period it featured in the DSM III), for we are not supposed to blame the mentally ill (as the common view holds). Odd as it may sound, freeing homosexuality from the list of mental illnesses represents a restoration to potentially be morally judged and thus, progress from viewing homosexuals as diseased agents in need of a cure. In this case, removing mental illness as a reactive attitude blocker represents a restoration of agency to the class of homosexual individuals. I myself find the moral judgment of homosexuality indefensible. Nonetheless, I find the diseasing of homosexuality even worse. (There obviously remains a need for further progress in reducing negative moral judgment of homosexuals. We should have never included homosexuality in the DSM in the first place and we should likewise exclude it as grounds for moral condemnation.

And in many respects, religion and science alike have treated homosexuals with similar indifference with respect their agency. So, I suppose ‘progress’ here would simply mean correcting one part of a larger error.) Now, I suggest we may be faced with a similar decision with respect to the psychopath. Perhaps we should acknowledge that this is a disorder that, like homosexuality, never belonged in the first place and that individuals so labeled need to be, once again, viewed as moral agents and with all the respect that this so accords, including unfavorable and even unfair moral judgments. For, it is better to be morally condemned, harshly or even inappropriately, than it is to be de-humanized and to be viewed as a non-agent unworthy of inherently human responses. If this is so, then we should resent the psychopath rather than see him as a non-agent or worse, a beast. It is hard to overestimate how important this distinction is when discussing the responsibility of the psychopath and how we should respond to him. So part of the disagreement about the responsibility of psychopaths may then have something to do with an underlying disagreement as to whether we find it better to pathologize persons while refraining from moral assessment altogether or, alternatively, whether we find it an improvement of some kind that we refrain from pathologizing persons whilst then engaging in moral assessments of them. I think the clinical history of homosexuality shows us the advantages of the former over the latter.<sup>11</sup>

---

<sup>11</sup> This is precisely Morris’ (1968) argument that there exists a right to punishment, as punishment is generally and morally preferable to involuntary psychiatric treatment. Deigh (1996) dissents from this view.

Let us now turn to examine the actual diagnostic criteria for Antisocial Personality Disorder where we will see one of the most overtly value-laden mental illnesses.

### **1.3 The empirical argument**

In the introduction to this dissertation I claimed that the question of whether or not to 'hold' psychopaths responsible (including the question of *how* we might do so) has been overly subordinated to the question of whether or not the psychopath 'is responsible' for his actions, omissions, and the resulting consequences. Levy (2007b) is explicit that when we take up the question of psychopath's responsibility that we are engaging in a factual inquiry (as opposed to a normative one) rather than an 'as if' inquiry: '[But] and agent's moral responsibility, or lack thereof, is not something we can simply *decide*. We are not here concerned with whether it is appropriate to act *as if* psychopaths were responsible. We are instead concerned with a *factual* question: whether they *are* in fact responsible for their actions,' (Levy, 2007a).<sup>12</sup> So, on Levy's view, we can *only* provide an answer the 'hold responsible' question(s) (whether or not we should, and if so, *how* we should hold psychopaths responsible), after settling on an answer to the (allegedly) factual question whether or not the psychopath 'is in fact' responsible. I am skeptical that our responses to psychopaths are afforded the luxury of factually establishing the moral responsibility of psychopaths beforehand. Below I engage the empirical data that

---

<sup>12</sup> Levy (2007a) p. 137, fn. 2

Levy uses to support the conclusion that psychopaths are not responsible directly. But before doing so, I wish to note two concerns with Levy's approach.

First, Levy assumes that the very concept of moral responsibility is an empirical, non-normative concept in which there appears very little, if any, middle ground. Thus, psychopaths either are or are not responsible on his view. This, however, is an overly bifurcated view of responsible agency. We are better served (theoretically and practically) by assuming that responsible agency can be better understood as falling along a continuum (the cases below/above provide some intuitive content to show this). We want to know where psychopaths fall along the continuum of responsible agency and whether or not they are destined to maintain a fixed position along that continuum. Evidence suggests that psychopaths, slowly but naturally, advance along the responsibility continuum (evidence that Levy ignores).<sup>13</sup> (Later in the dissertation I return to discuss whether or not a reduction of symptoms can be construed as an increase in moral responsibility. I argue that it can.) Furthermore, the very exercise of positioning various agents along the responsibility continuum is a *normative* exercise. Granted it is an exercise that should take empirical data into consideration – but only on the understanding that the evidence is provisional and that considering the evidence does not commit one to a value-free understanding of *either* responsibility *or* psychopathy.

This brings me to the second concern, that Levy's view appears to commit us to *not* develop a policy regarding how we should hold psychopaths responsible until we have established the fact of psychopathic (non)responsibility. Scientific

---

<sup>13</sup> DSM IV TR

reasoning remains in flux, as does the gathering of empirical facts. Otto Neurath once described the process of gaining scientific knowledge akin to building a boat while at sea without ever being able to dock.<sup>14</sup> So, the establishment of the neural basis for moral responsibility, even if taken as a strictly empirical enterprise, is an inherently speculative endeavor. Nor is it clear that the concept of responsibility is as thoroughly empirical as Levy suggests. Thus, we are *forced* to provide a response to the normative question of how, and whether or not we should go about holding psychopaths responsible absent any factual certainty regarding his responsibility. Levy too is thereby committed to responding to the 'hold responsible' question with only a (weakly) suggestive gathering of evidence to the allegedly established 'fact' of psychopathic moral responsibility. Before criticizing what I take to be an inversion of appropriate normative responses to psychopathic wrongdoing that Levy calls for (by default), I wish to take a closer look at the evidence he offers to the conclusion that psychopaths are, factually speaking, not morally responsible.

#### **1.4 No escaping values: criticism of the empirical argument**

I can only here provide a brief summary of the neuro-empirical evidence provided by Levy in support of the conclusion that psychopaths are not morally responsible, as a more in depth discussion would steer this dissertation too far into the domain of philosophy of science. (Although it should be noted that the

---

<sup>14</sup> Cartwright, Cat, Fleck, Uebel (2008), p. 89; Neurath (1932), 'Protocol Statements', *Erkenntnis* 3, p. 204-14.

philosophy of psychiatry requires that we engage in philosophy of scientific reasoning every bit as much as it requires we engage in moral and political philosophy.) After summarizing Levy's account of the empirical evidence, I raise two main objections: first, that the evidence falls considerably shy of being suggestive, and second, perhaps more importantly, that Levy's own view is inescapably evaluative and that he appears unaware that his view is so committed.

The basic empirical approach to explaining psychopathic (non) responsibility includes brain scan imagery by fMRI of psychopathic subjects, while slightly more complex approaches include imagery of psychopath's brains while being asked various moral thought experiments, (for instance, variations of the trolley problem). Two regions of the brain have been identified as playing a crucial role in our ability to think and behave morally, the amygdala and the ventromedial prefrontal cortex. Initial research (Blair, 2005) suggests that the amygdala, the region of the brain thought to regulate emotion, is impaired, or relatively smaller in psychopathic brains than in the brain of non-psychopaths. However, amygdala dysfunction is believed to account for the *reactive aggression* that is associated with some psychopaths but it does not account for *instrumental aggression* associated with other psychopaths. Instrumental aggression is distinct from reactive aggression in that the former lacks the loss of emotional control that defines the latter. Instrumental aggression requires a cool head and an ability to put together a complex plan of action leading to a specific desired result (often the assault or killing of the target of the instrumental aggression). As noted by Levy (2007a), it is the 'deliberate use of instrumental violence that makes psychopaths so hard to

forgive.’ Furthermore, the level of agency required to deliberately engage in instrumental violence belies the (Strawsonian) assumption that psychopaths obviously lack agency.

An alternative neuro-cognitive account places the seat of responsibility in the ventromedial prefrontal cortex (VM), as patients with VM damage tend to (mis)behave similarly to psychopaths., (Roskies, 2003). (There is a corollary debate whether or not VM patients are morally responsible that has some relevance to the question of psychopathic responsibility but I will ignore that debate in this dissertation.) The point of relevance here is that Levy, assuming VM patients are like psychopaths, appeals to the performance of (and brain imagery of) VM patients when asked about moral dilemmas like variations of the trolley problem as providing empirical (hence, value-free) evidence that VM patients are not morally responsible. As Levy explains (2010, p. 217), that VM patients ‘perform normally’ when asked to flip a switch that redirects the trolley to hit one worker instead of five, but when asked whether to push a bystander in front of the trolley to redirect it to hit one worker instead of five, VM patients ‘perform *abnormally*,’ [emphasis added]. On Levy’s view, the VM patient does not answer *wrongly*, or even *controversially*, he answers *abnormally*, because of his neural pathology.

Philosophers are familiar with the trolley problem in part as a tool for distinguishing utilitarian (welfare maximizing) approaches with deontological approaches, and while we philosophers might have reasonable disagreement about what to do in various trolley scenarios, what we do *not* do is dismiss the view of our counterpart as abnormal or brain damaged, or otherwise pathological. Levy here

appears to pathologize a specific moral judgment.<sup>15</sup> (Again, the *abnormal* judgment here is attributed to the neural damage; hence its abnormality is pathological.) In as much as we might disagree with many utilitarian judgments (or the theory more generally) we do not dismiss such judgments as evidence of pathological moral incompetence. Fortunately, there is evidence (overlooked by Levy) to bolster my point. The research of Cushman and Young (2009) provides evidence from non-VM damaged, non-psychopathic ‘normal’ individuals in which different brain regions under fMRI reveal signal changes in the brain that correlate with different moral theories held by those individuals. Briefly, the ventromedial prefrontal cortex seems to be correlated with conforming to deontological rules, (the VM ‘illuminates’ on fMRI when deontologists say ‘no’ to pushing the bystander onto the track) while an idle ventromedial prefrontal cortex (VM fails to illuminate) is correlated with welfare maximizing decisions (pushing the bystander onto the track). Since two distinct types of moral reasoning occur in two distinct regions of the brain, (only one of which is associated with the region of the brain that Levy believes to be involved in moral reasoning), what are we to make of this? Either we conclude that localization theories about moral reasoning are too nascent in their development to tell us very much about responsibility, or we can conclude that utilitarians are psychopaths. Since utilitarians are not psychopaths, (despite similarities of brain structure when making moral judgments), then the state of the empirical evidence is

---

<sup>15</sup> Levy (2010); to fill out his argument here: VM patients – owing to neural damage similar to psychopaths - lack moral competence and are thus, not morally responsible. And by extension, utilitarians likewise lack competence and responsibility. This is of course, false.

far from compelling, and is perhaps not even suggestive in the slightest. More importantly, however, is that we note how Levy has pathologized one set of responses based on fMRI 'evidence'.

Hence, we can conclude that Levy's heavy reliance on new empirical data fails on two accounts. First, Levy's insistence that moral responsibility itself is an empirical matter (2007a, p. 169) that can be established without doing moral philosophy appears to be achievable only by simply redefining moral values (especially *bad* values) as pathological values. We should reject this move on the most basic Szaszean grounds that we should resist the 'medicalization of morals.' Second, we should likewise be weary of the smuggling, however overt, of values *into* the concept of psychopathy itself. What Levy takes as neuro-empirical evidence of *mad* values appears to be little more than evidence of *bad* values, (assuming that we reject utilitarian judgments on moral grounds and not on medical grounds).

We can agree with Levy that continued neuro-empirical research (by fMRI) of psychopaths is warranted but disagree with the implications this research will ultimately yield. Levy appears to think that improved empirical research, (say, that identifies with more explanatory authority the specific brain region that is faulty in psychopaths), will show that we should exempt psychopaths from moral responsibility. To the contrary, if we possess evidence with more explanatory power, perhaps we should rather use this evidence as a contribution to understanding *how* to help psychopaths meet even the most basic moral obligations (thereby, morally improving). Better evidence need not imply a better case for exemption. Rather, better evidence implies developing our approach to

psychopaths *qua* potentially more fully responsible agents. (Here our research might now ask, 'can we bypass the VM in moral reasoning?' and, 'how can we increase range of emotional responses of psychopaths in light of VM data?')

If in the future we discover the neural structure of moral responsibility, then we still have a choice of what to do with this knowledge. On the one hand we might conclude that psychopaths are indeed exempt from responsibility and decide from this premise (now a factual one, we are assuming) what to do with psychopaths. Perhaps, following the common view, all offending psychopaths are to be detained indefinitely, but are never to be blamed. Perhaps we develop a screening device that allows us to preemptively round up and detain psychopaths even prior to their offending, (justified by the statistical likelihood that they will assault or kill someone), but we are never to blame them. At this point a dissonant intuition needs to be identified, that something seems morally wrong with rounding up and jailing psychopaths whom have never harmed anyone or otherwise offended. We would be preemptively jailing the non-offending psychopath based on a neural profile that predicts the psychopath will inevitably assault or kill. I assume without argument that it would of course be immoral and unjust to preemptively detain *anyone*, including psychopaths. Perhaps this assumption does not extend to psychopaths and we would want to know the predictive accuracy the neural profile provides, and if it is high enough (say 99% likelihood) then we might be justified in detaining all psychopaths, even the 1% for whom we are not certain will offend. But this would be a justification that relies on utilitarianism, and thus, is not likely one Levy would find sufficiently *normal* to support. If Levy, owing to his own deontological bent,

would reject the preemptive detention approach, then he must admit that what still matters the most (more than a neural profile) is *what the psychopath in fact does*. And, to own the hidden premise here, I presume that *if* we are assuming a natural kinds model of psychopathy in which the psychopath is defined by his uniquely abnormal neural structure, (a view Levy appears steadfastly committed to), then we can also assume that there are psychopaths who refrain from acting wrongly. In fact, at least one such case has been documented, in which a neuroscientist studying brain images of psychopaths learned after looking at his own brain that his neural profile matched that of psychopaths.<sup>16</sup> Yet this man has no criminal record, is involved in normal interpersonal relationships (is married with children, with professional colleagues, etc.) and certainly does not engage in cool, calculated instrumental aggression. Even if we assume there is a neural basis to psychopathy, it does not follow that all agents in the neural class will go on to aggressively harm others. Of course, and this point cannot be under-emphasized, *there is no such thing as a non-offending psychopath*. In order to ‘earn’ the diagnosis, one has to have actually exhibited a habit of harming others. So, psychopathy is not so much in one’s neural structure as it is a pattern of behavior. Furthermore, possessing the identical neural structure of psychopathy appears not to be as causally deterministic.

Taking stock of the dialectical situation at the close of this section, I have argued that the empirical evidence cited by Levy that alleges to explain why psychopaths are not responsible does not pass muster. In fact, Levy’s treatment of the evidence reveals a crucial error that, swiftly and without argument, simply

---

<sup>16</sup> Hagerty (2010). The neuroscientist’s name is James Fallon.

redefines controversial moral judgments (utilitarian judgments specifically) as *abnormal* and pathological judgments. Put in the most simple of terms, Levy has relabeled *bad* moral reasons as *mad* moral *unreasons*. Furthermore, from this paradigmatic shift, it is then assumed that since mad reasons are *caused* in a way that bad reasons are not, and since psychopaths can only have mad reasons, then they do not possess the causal control over their reasons in a way that allegedly grounds responsibility. But as we have just seen, once the bad to mad mistake made by Levy is revealed, then the causal basis for psychopathic ‘reasons’ for acting loses its intuitive metaphysical punch. It also shows that we cannot simply dismiss the reasons psychopaths offer for their behavior as causes and not reasons at all. We must confront the reasons and however unreasonable they are we should confront them as belonging to the normative domain of reasons. Later in chapter 4 I defend the targeting of reactive attitudes toward psychopath’s behavior.

### **1.5 Antisocial Personality Disorder: value-laden diagnostic criteria**

There are several criteria that one must meet to earn a diagnosis of Antisocial Personality Disorder. (It should be noted that the period of writing of this dissertation straddles two versions of the DSM, the revised fourth edition and the very recently published fifth edition. I focus exclusively on the fourth edition for two main reasons. First, Levy provides the most cogent argument to excuse the psychopath from responsibility and his essays were written and published during the time of the fourth edition of the DSM, and second, because a lengthy discussion

of the differences between Antisocial Personality Disorder, minor as they are with respect to the presence of values, would take the discussion too far adrift. I will say only briefly that the DSM V represents an entrenchment in a value-laden conception of *all* so-called Personality Disorders, including Antisocial Personality Disorder. The notable shift in language to emphasize ‘pathological personality’ as the core concept sadly ignores a debate previously discussed at length that ‘function’ is not the value-free Holy Grail it has been argued to be (Boorse, Wakefield, Fulford). The DSM V then, unfortunately, represents a steadfast entrenchment to an alleged value-free conception of, at least, the Personality Disorders.)

Criterion A and its seven sub-specifications are of particular interest to our analysis. (B, C and D are ‘differential diagnosis’ criteria, which are designed to distinguish Antisocial Personality Disorder from other mental illnesses.) The task, then, is to determine whether or not values are present in these criteria, and if so, what type of values are involved. By showing that specifically *moral* values are involved in the criteria, and thus, conception of psychopathy, we can see that the task of diagnosing and *identifying* psychopaths is normative. And if psychopathy can indeed be shown to be conceptually normative in content (that is, for diagnosis), then the argument that psychopathy is a value-free disease that independently causes immoral actions, relies on an erroneous assumption that psychopathy is a disease. As we shall see, the diagnostic criteria of Antisocial Personality Disorder, are not based on histology and require no identification based on physiology. This observation is particularly pertinent to the forthcoming discussion of Levy’s

speculation that the bad physiology of psychopaths explains their immoral, or rather, pathological behavior.

The pertinent criteria for diagnosing psychopathy lie in the seven patterns of behavior, of which only three need to be exhibited to earn the diagnosis. The identification and diagnosis of Antisocial Personality Disorder is determined by the following (from DSM-IV, 1994):

Criterion A: There is a pervasive pattern of disregard for and violation of the rights of others occurring since age 15 years, as indicated by three (or more) of the following: having hurt, mistreated, or stolen from another.

1. Failure to conform to social norms with respect to lawful behaviors as indicated by repeatedly performing acts that are grounds for arrest.
2. Deceitfulness, as indicated by repeated lying, use of aliases or conning others for personal profit or pleasure.
3. Impulsivity or failure to plan ahead.
4. Irritability and aggressiveness, as indicated by repeated physical fights or assaults.
5. Reckless disregard for safety of self or others.
6. Consistent irresponsibility, as indicated by repeated failure to sustain consistent work behavior or honor financial obligations.
7. Lack of remorse, as indicated by being indifferent to or rationalizing.

Notice the absence of etiology as a diagnostic requirement. The DSM (any edition) does not require any identification – or even speculation - of the cause of

psychopathy in the diagnosing the disorder. Others however, particularly Levy, have much to say about potential causes of psychopathy, including how etiology informs the question of the responsibility of psychopaths. The DSM does not say why or how this behavior is brought about in such individuals. It only requires *that* such patterned behavior be evident for the diagnosis to be met.

Although reference to social and legal norms is explicit in criteria 1 and 6, the DSM does not directly address moral responsibility. Moral concepts are of course deeply embedded in the language above, specifically in criteria 2, 4, 5 and 7. The meta-criterion, a 'persistent pattern of disregard for and violation of the rights of others', is a phrase issued in moral language that is particularly noteworthy. Since 'rights' as used here (in meta-criterion A) is not parsed along social, political, and moral lines, does not mean that moral rights are excluded. Rather, there seems no reason not to assume that moral rights are included, despite the rather wide sense in which 'rights' is used here. Thus, the diagnosis and concept of psychopathy cannot purge moral concepts from its content.

This requires that we acknowledge that any view of the responsibility of psychopaths cannot simply operate under the assumption that psychopathy is a medical concept that seeks to be, if not value-free, as value-purged as is possible. However, as we see throughout the diagnostic criteria, psychopathy is saturated with values. It is hard to even imagine what would remain, and in strictly medical terms, if all values were indeed purged from the criteria for and concept of psychopathy. The source of tension (and dispute) with respect to the responsibility of psychopaths stems from our commitment to either the notion that psychopathy is

a medical concept or, alternatively, a moral concept. Must we resolve this dispute before reaching a conclusion about the moral responsibility of psychopaths?

Perhaps we do not need to resolve this issue (that is too much to ask) but neither can we ignore it. We should resist the impulse to excuse or exempt others from responsibility simply because of the presence of a medical disorder. Of course, as the criteria above show, it is wrong to think of psychopathy as a medical disorder.

The tendency to sharply divide medical disorders as inappropriate targets of reactive attitudes like blame and resentment from non-medical problems in living as, somehow, appropriate targets of the reactive attitudes leaves us with two equally stark options with respect to the responsibility of psychopaths. If we see psychopathy in purely medical concept, then he is not responsible at all.

Alternatively, if we see psychopathy as a value-laden concept, then he is at least somewhat morally responsible. Additionally, we might wonder why psychopathy *as* a medical concept was ever conceived to begin with. Rather than categorizing psychopathy as a medical disorder that alleges to explain the lack of moral capacity, perhaps psychopaths are more like racists and misogynists in which the lack of moral capacity is does viewed as a medical ailment and is subject to the moral judgment of others. Psychopathy appears to be a paradigmatic case of what Szasz (1974) called the 'medicalization of morals.'

My main aim here is to undermine two incorrect assumptions (held by the common view) that work in tandem. First, I have objected to the conception of psychopathy as a medical disorder, and second, that this alleged medical disorders (even a disorder that 'targets' a persons moral responsibility) causes its victims to

be not responsible, thus making it inappropriate to hold them responsible for their wrongdoing. This view of the psychopath, (what I call for brevity the 'common view'), is implicitly held by Strawson and explicitly argued for by Levy.

Can we infer the presence of a medical disorder from observing behavior that routinely deviates from social norms? A key source of confusion in the debate over the responsibility of psychopaths lies in the fact that paradigmatic psychopathic behavior is exactly the same as paradigmatically immoral behavior. Since psychopaths routinely engage in these behaviors (that are, at once, immoral and psychopathic) then we are left to infer one of two things regarding their responsibility: either psychopaths are the most immoral among humans, or psychopaths are the least immoral among humans. These are two rather stark and polarized options.

The psychopath apologist appears to assume that from the unusual regularity of immoral behavior that we can infer moral blindness. This of course does not follow. We might just as well infer that persons whom are so routinely engaging in immoral behavior are not morally blind but rather place no weight on many of the basic moral commitments made by others. So, rather than inferring that psychopaths are morally blind, as is commonplace among philosophers, we might alternatively infer that they willfully dismiss the basic moral demands (say, refraining from assault, murder and rape). Here then we have a 'seeing as' problem. The psychopath apologist sees that psychopaths *cannot* see that the interests of others should factor into his behavior, whereas others see that psychopaths simply

*do not* consider the interests of others. The difficulty in providing an account for either case is that the inferences we make are *based on behavior*.

If psychopaths are indeed responsible, then, owing to the severity and regularity of their morally offensive behavior, then they are the *most* blameworthy among moral wrongdoers. Psychopaths are *serial wrongdoers* whose wrongdoing includes violence and psychological and emotional manipulation of their victims. But, this is precisely what the common view encourages us to reject. The psychopathy itself is alleged by the common view to cause psychopaths behavior in a way that exempts him from moral responsibility. So, according to the common view, to call psychopaths 'serial wrongdoers' is a misnomer or is at the very least, question begging. I hasten to emphasize, again, that the criteria required to diagnose a psychopath are written in behavioral terms that *include* negative evaluations: 'persistent disregard,' 'reckless disregard' 'consistent irresponsibility,' 'mistreat[ing],' and 'indifference' to wrongdoing. A diagnosis does not rely on nor refer to causal explanations; and certainly does not include any histological claims. So, the absence of moral capacities exhibited by psychopaths may be more like those of the racist and misogynist than like that of an animal or winter storm.

Suppose for a moment the common view is correct and that psychopaths, owing to the disease over which they have neither control nor responsibility in contracting, are not morally responsible for their 'wrongdoing.' If so, then what about the moral status of individuals who fall just short of the clinical threshold for psychopathy? Would these then be the most blameworthy of moral wrongdoers (albeit with slightly less persistent wrongdoing than the psychopath)? Call these the

'subclinical psychopaths'. How might we morally respond to these folks? Now, further imagine that a person comes to learn that their behavior falls just shy of clinical psychopathy. Should subclinical psychopaths intentionally add more regularity to their wrongdoing to meet the clinical threshold that they *ipso facto* become *unblameworthy*? This seems odd.

Discussions of the responsibility of psychopaths frequently ignore the fact that psychopaths are defined by behavioral traits and also overlook the fact that moral conduct requires a kind of skill development (part social, part cognitive and part emotional). Apparently, psychopaths did not learn these skills. But can they be taught the skills? We don't know the answer to this question because we (as an intellectual community) haven't invested very much at all in the way of resources towards a better-informed answer. And so we need not preempt the question of engaging the psychopath with a posture toward his moral improvement.<sup>17</sup>

We are happy to teach our children to develop moral skills. In fact, much of parenting is devoted to this very task. But for some reason we think that only children are worthy of such investment and that by the time one reaches adulthood one has aged out of such development. What this translates into for adult psychopaths is something like the view that 'its just too late' to teach them morally.

---

<sup>17</sup> A related set of moral questions arises in this context that I will not discuss in this dissertation – perhaps only in the conclusion – include: should moral skill development be the domain of clinicians and therapists? If so, this marks a paradigmatic shift in the place of moral values in the therapeutic milieu. Also, beyond talk therapy, what other specific interventions are morally justified? Last, what interventions are justified under the assumption that psychopaths will not voluntarily consent to treatment? How may we morally justify manipulating psychopaths into treatment designed to produce moral improvement?

This is a key assumption of the common view and a major problem in our discussion of the responsibility of psychopaths, including how we should respond to them. I suspect that if we can move beyond the idea that there is a developmental terminus with respect to age and moral development, that we can make space for the idea that psychopaths, however delayed they are morally, can learn to behave with increasingly moral consideration. If this is plausible, and I think it is, then we may be obligated to respond to psychopaths with consideration for their moral development. But here we meet another problem and that is the assumption, held in the common view, that the reactive attitudes do not work that way. The reactive attitudes, we are told, are supposed to align with a clear answer in mind to the offending agent whom either is or is not obviously a moral agent. As we have previously discussed in the cases above, unlike many other severe mental illnesses, psychopaths yield a strong intuition that we are dealing with an agent ensconced in reality (unlike the cases of Harold or Lori discussed below). Strawson believes that it is so intuitively obvious that psychopaths lack moral agency (and that we should adjust our reactive attitudes accordingly) that stipulating the 'moral idiot' as beyond any and all moral understanding comes without question. While I generally agree with Strawson that we need to attune our reactive attitudes toward psychopaths, I disagree with Strawson as to why and how we should tailor our moral responses to wrongdoing psychopaths.

This leads to an important epidemiological observation about psychopaths. As with many of the other personality disorders identified in the DSM, individuals diagnosed with Antisocial Personality Disorder experience a pronounced decrease

in symptoms beginning around age 40.<sup>18</sup> We don't know why this occurs but epidemiological data bears out this observation. This is a key observation in our assessment not only of the responsibility of psychopaths, but in formulating responses to their wrongdoing, including whether or not we should hold them responsible and if so, *how* we should hold them responsible.

At this point we need not assent to the claim that, since we will never gain the quality of response from the psychopath that we desire and often receive from wrongdoing non-psychopaths, this is somehow evidence that psychopaths will *never* be able to increase moral responsibility by better meeting basic moral expectations. The lack of moral responsiveness so regularly displayed by the psychopath need not be confused with an incorrigible lack of moral responsibility. Psychopaths are notoriously difficult to clinically 'treat', however, and our collective lack of success leads many to the erroneous conclusion that they are incorrigibly and fully morally blind. The lack of success in 'treating' psychopathy, however, could just as well be a commentary on the current state of the art rather than evidence of psychopathic incorrigibility.

## **1.6 Cases**

To chart intuitions regarding the relationship between mental illness and moral responsibility, I offer several cases (some inspired by real people and some imagined) meant to test our intuitions about responsibility and mental illness.

---

<sup>18</sup> *DSM IV-R*, Axis II Personality Disorders, Antisocial Personality Disorder: Pervasiveness and Treatment

Examining a variety of cases allows us too situate the psychopath among some other mental illnesses in order to highlight the particular difficulty involved in assessing the responsibility of the psychopath. I first present the cases in a bunch and then follow with discussion of each case as well as the collective cases.

### **Harold in the waiting room**

While awaiting an appointment with his psychiatrist, Harold suffers from a psychotic episode that includes visual hallucinations of being attacked by flying body parts. In an effort of self-defense he throws furniture at the non-existent assailants, leaving the waiting room in wreckage. Fortunately, others seated in the waiting room were able to escape with only minor injuries.

### **Lori**

Cousin Lori has severe cognitive impairment (an IQ of 50) that leaves her unable to communicate verbally, with very limited sign language ability. Lori is physically strong and she appears to enjoy physical embrace with people she recognizes as familiar. There are times when Lori's endearing embrace of a child she recognizes results in physical harm including scratches and bruising.

### **Pedophilic Disorder**

Imagine an abstract agent diagnosed with Pedophilic Disorder. He was sexually abused as a child and as an adult he repeats the offense.

## **Robert Harris**

The case of Robert Harris has received attention from philosophers following Gary Watson's (1993) introduction of him. Harris was a young man with a violent and criminal history that culminated in the murder of two teenage boys, a crime for which he is eventually executed. Harris had previously beaten another person to death but was not punished (for reasons unclear). Harris, we also learn, suffered a childhood filled with profound physical and emotional abuse. As his execution drew near, Harris confessed that, "[he] had the chance and took the road to hell and there's nothing more to say," (Watson, 1993, fn. 22).

## **Tom**

Tom has an extensive garden-variety criminal history that includes, notably, theft, extortion and numerous instances of physical assault, (with women as his predominant target) and sexual assault, (primarily of men and young men, offenses for which he was legally charged but which he voluntarily revealed nonetheless). Tom is boastful of his criminal past and appears to take particular pride in boasting about his pugilistic skill, which he credits alongside his 'street smarts' for making his way in the world. Tom is highly verbal and has a slightly higher than average I.Q. Owing to Tom's convincing and sincere performances before parole boards, previous sentences have been reduced when, swayed by his optimistic future plans and his own moral rebirth he is released early from prison. Tom is serving a second life sentence for beating his most recent wife to death. When asked what led to the beating Tom replies, 'she said I was fat and she had to be taught a lesson.'

## **Ella the Empath**

Ella is the antithesis to the psychopath. Rather than having no moral sense at all, Ella has too much. Ella has been diagnosed as an Empath (according to the future DSM VI), which is a disorder of personality marked by overwhelming moral sentiment toward others in such a way that affects her social and occupational functioning and is, hence, pathological.

### **1.7 Discussion of Cases**

In this section I provide a general discussion and canvassing of intuitive responses to each case with these questions in mind: What bearing does the mental illness involved in each case have on the responsibility status of the individual? Are some of these individuals more responsible than others? If so, does the specific mental illness involved *make* an intuitive difference in the degree of responsibility of these individuals? Later in the dissertation I discuss how these questions inform our moral response to these individuals, with detailed focus on the psychopath.

#### **Harold in the waiting room** [Active psychotic phase]

Harold has been diagnosed with a psychotic disorder. Given that Harold is hallucinating and that his aggressive furniture throwing is directed at objects that do not exist (at least not to anyone else), there is a strong intuition that the minor injuries that he causes others are at the very least, not intentional. (We must also be careful here to distinguish hallucinations from delusions; while both are symptoms of psychosis, the former are far more severe and rare occurrences than the latter.)

Since Harold does not intend to harm anyone, and only intends to immediately defend himself from perceived attack, many will find that he is not responsible for the harm he causes. Here, assessing Harold's responsibility focuses on the nature of his intentions and how they are affected by the hallucinations he experiences. The absence of the intention to harm (since it's hard to tell if Harold is even aware of the presence of others) intuitively diminishes the impulse to hold him responsible, say by verbally blaming him for committing a moral offense. Although there is no direct intention to harm, is Harold nonetheless responsible for the consequences of his actions? Harold's actions do result in (albeit minor) harm to others and, in some sense, he causes the harm as well as destruction of physical property, but it is inaccurate to say that he intended to bring about these consequences. His intentions contained no calculation of consequences. We might, however, ask Harold to apologize to the harmed bystanders even though we recognize that he did not intend to harm anyone. Perhaps we should stick Harold with a bill for damages. But placing a financial debt on Harold may be overly censorious and harsh, especially if he is without financial resources to pay. If, however, we place a moral psychological debt on Harold, say, by asking him to apologize, are we likewise asking him to pay a moral debt even if he does not have the resources to pay? If so, we would be putting Harold in a kind of moral debtors prison and this is not a conclusion we should find satisfactory.

We should also recognize that when Harold emerges from the active phase of his psychosis and recognizes that he caused damage and harm, even though he was unaware of doing so at the time, that he may very well reach the conclusion,

unprompted by a third party, to apologize for the damage and harm. Later, we will see that Harold's case aligns with what I call Strawson's type 2a case, in which our reactive attitudes *toward an agent* is temporarily suspended and this temporary suspension aligns with Harold's temporary suspension of agency (a suspension of agency that is defined in part by an 'innocence' of the intentions of his actions). What does not follow in Harold's case is that we suspend reactive attitudes toward him *for all time*, because his agency is not suspended for all time. Thus, we tailor our moral demands, including our holding him responsible, around the ebb and flow of his agency that is, of course, affected by the tides of his psychosis.

We need not determine a more specific response to Harold here. What is more immediately pertinent to our discussion is the way in which the hallucinations and his mental illness erode the intention to harm others and the impact this has on our understanding of his responsibility. Since Harold's actions are a response to hallucinations it is safe to assume that his disorientation is global, that is, he is not oriented to circumstance (this phrase is less vague than say, 'not living in reality'), including not being oriented to the presence of other people in the room. Lacking orientation to circumstance and place (perhaps even lacking orientation to time and person) suggests that his mental status is comprehensively diminished. Whether the hallucinations arose as a side effect from a toxic cocktail of psychotropic medicine or perhaps arose directly from his psychosis (though, unlike delusions, psychosis induced hallucinations are very rare) the role they play in his actions, rather his reactions to them, diminish the intent of his behavior so globally that moral intent washes out alongside all other intent, aside from his basic intention to

defend himself. Perhaps it is this vulnerability of Harold's condition that diminishes reactive attitudes that we might otherwise hold or direct toward him.

Even though Harold has been diagnosed with a psychotic disorder and he is in the throes of its active phase, we need not reach the conclusion that he is fully exempt from moral responsibility such that we disengage from him. Rather, an informed response to Harold following the waiting room fiasco is to attend to his agency, including his moral agency, with diachronic considerations. We might find that in time we discourage Harold from continuing to feel remorseful for the waiting room fiasco. In doing so we would be discouraging him from lingering on a self-related reactive attitude of blame. Compassion for Harold leads us to diminish third party blame and even to encourage a reduction in his own self-blaming.

We cannot, however, generalize from this conclusion that there are no cases involving mental illness in which holding the person responsible (including expressing blame) is not likewise motivated from compassion. As we will see the case of the psychopath shows this.

**Lori** [Severe to Profound Cognitive Impairment, formerly termed Mental Retardation]

Lori has Down's syndrome and an IQ of 50 which she has maintained throughout her adult life, hence her diagnosis of Severe Cognitive Impairment. (Revisions to the DSM V entitle the diagnosis as an Intellectual Developmental Disability or simple, Intellectual Disability). So, unlike Harold, Lori has no fluctuation of agency. Despite the efforts of those involved in her life (teaching her

adaptive skills and sign language with some success) her cognitive development is static and has long since reached maximum threshold. Owing to her reduced cognitive state, Lori is situated, if not at the far ends of the Not Responsible continuum, then very near the far end. Lori's placement at this end of the continuum results from a global and comprehensive cognitive deficiency.

How should we respond to Lori when she scratches her cousins while embracing them? At the time of the scratching embrace she appears to others as having no insight or awareness that she is causing harm. There is no malice of intent involved and the enthusiasm of her embrace appears gives the distinct impression that she enjoys it. After all, she initiates the embrace. But she appears unaware of the harm she is causing. Her reduced cognitive state does not mean that we cannot increase her awareness that she scratches when she embraces and, in turn, reduce the occurrence of scratching. If we can in fact alter her behavior, we should. Furthermore, we should diligently seek out any technique or intervention to bring about this result. Now, to put squarely in our focus the question of holding Lori morally responsible, should expressions of blame or resentment be included among these interventions?

Given that we have already tied her comprehensive cognitive deficiency as good reason to place her on the Not Responsible end of the responsibility continuum, then it seems to just follow that holding Lori morally responsible is pointless at best, and at worst, grotesque. Why is this? And just what would it mean to hold Lori morally responsible? Our intuitions regarding the hold responsible question with respect to Lori remain dormant, preempted by our negative response

to the 'is responsible' question. Since she is not responsible, then it follows that we should not hold her responsible. Although the ordering of these two considerations (the 'is responsible' and the 'hold responsible') seems accurate in Lori's case, this pattern need not be viewed as the standard by which we should consider all other mental illnesses with respect to the question of responsibility.

The question I asked above seems additionally confusing by asking if (maybe even presupposing) reactive attitudes and their verbal expression (what we typically take to mean 'holding responsible') are properly conceived as 'techniques' or 'interventions.' One might object that the language of 'intervention' is inherently clinical and that to hold someone morally responsible is not simply a matter of behavioral modification. But I think this objection is misguided, for there do seem to be appropriate circumstances in which our holding someone responsible contains the goal of behavioral modification by way of moral instruction. With respect to Lori, the behavioral instruction we provide need not be inherently moral in its content (in fact, I'm not even clear what the phrase 'inherent moral content' means). But the behavioral modification we seek may very well have the morally worthy motive of seeking ways to keep Lori in touch (physically and psychologically) with other humans. What is it then, in this response to Lori, that is missing from the moral equation? It seems, only a commitment to the notion of moral inherency (a commitment whose time has perhaps passed) and the confounding confusion created by assuming that an attitude of resentment (expressed by verbal shaming) somehow provides evidence of this moral inherency. Lori is better served if we hold her responsible by way of placing behavioral expectations on her and providing her

with the instructional support and resources to meet those expectations. Perhaps we simply take more caution that her fingernails are clipped and maintained at a length that prevents scratching in the first place. This seems to be a superior and more humane response than say, isolating her and forbidding her to embrace others.

Even Lori's case, and like persons with severe cognitive impairment, for whom there is likely widespread consensus that she is not morally responsible, shows that there exists asymmetry with respect to the is responsible question and the hold responsible question. Our response to Lori's scratching behavior includes behavioral demands and expectations, and at least in some sense, perhaps the most important sense because it decreases her social isolation and rather seeks to maintain human contact with her, these demands are a manner of holding her responsible.

Many will find this notion of holding responsible a diluted notion, devoid of specifically *moral* content, and that modifying Lori's behavior (including placing behavioral expectations on her) is not what we mean when we speak of holding person's *morally* responsible. I can only respond to this worry briefly here but will discuss what I take to be an overly coarse distinction in more detail in chapter 2 when discussing the Strawsonian framework of reactive attitudes and the objective view. The short reply is that even a thoroughgoing, or 'pure' notion of holding someone specifically morally responsible cannot eliminate the underlying call for behavioral modification of the wrongdoer. Furthermore, should one person morally blame another (the 'personal reactive attitudes' in Strawson's terms) for her actions,

omissions, etc., *without* an underlying call for behavioral modification, that such a version of blaming is no ideal to celebrate and perhaps even a practice we should forego.

### **Pedophilic Disorder**

The short of the pedophile's story is that he is a serial child rapist and has been diagnosed with Pedophilic Disorder. We might imagine the journey of intuitions about the pedophile's moral responsibility to run as follows. Upon hearing that a given pedophile is a serial child rapist our initial attitude toward him is, to condense into one phrase, one of moral disgust. Rape is morally deplorable, even more so when children are the victims. But this initial presenting attitude may well be mitigated for some upon learning two things: first, Pedophilia is a diagnosable mental illness, and second, that pedophiles themselves are often sexually abused as children. Call this a post-initial attitude. Others may maintain the initial attitude of moral disgust even after learning this new information about the pedophile's condition and history. The divergence of post-initial attitudes renders the pedophile's case different from that of Harold and Lori, where no such divergence arises. Why the divergence in the case of pedophiles?

One reason there is no intuitive divergence in either the case of Harold or Lori is that the harm to others is so minimal that it is a stretch to say that anyone was victimized. Victimization affects our moral intuitions. It stokes them, in fact, and to many, the victimization that occurs in the pedophile's case eclipses what would otherwise be offsetting information upon learning of his disorder and

unfortunate history. Now, must our intuitions yield only two options here that the pedophile either is or is not morally responsible? I think not. One might possess competing intuitions, one that feels compassion and even moral sadness for the pedophile and his lost childhood *whilst simultaneously* feeling morally disgusted with him for the damage he is visiting on others. (I also suspect that philosophers who avail themselves of mental illness as a tool designed to show how cleanly we can divide morally responsible agents from the non-responsible ones avoid the example of pedophilia precisely because of the tension of intuitions that it yields.) The source of intuitive tension in the pedophile's case (should there exist any) shows that, despite the intuition that mental illness often serves as a responsibility blocker, this intuition can be offset by the presence of victimization. Furthermore, we can consider both competing intuitions about the pedophile's responsibility in forming our response to him.

It is worth noting that psychopaths, for all their alleged moral blindness, express outrage toward pedophiles. In so doing, psychopaths appear to be expressing a moral attitude, specifically, what Strawson calls 'vicarious reactive attitudes' that third parties feel on behalf of victims. This feature of a psychopath's moral psychology presents a powerful counterexample to the insistence that psychopaths are completely morally blind, or completely devoid of moral sense.

## **Psychopath cases**

### **Robert Harris and Tom**

This section combines discussion of both cases of Robert Harris and Tom. Watson (1993) introduced the case of Robert Harris to discussions of moral responsibility and its limits (never mind Watson's argument for the time being). The case of Tom is adapted from Hervey Cleckley's influential book on psychopathy, *The Mask of Sanity* (1955), in which Tom is one of many case studies.

Beginning with Harris, the flow of our intuitions might begin with an initial sense of horror and outrage when hearing about the particularly cold blooded, execution style murder of two teenage boys immediately after which he helped himself to the lunches of the victims as he laughed about his deed. If this is all we know of Harris' story we may well entrench this initial attitude of moral outrage toward Harris the adult. Upon learning about Harris' abusive childhood our moral intuitions divide. For some, the extent of abuse he experienced as a child serves to mitigate the initial attitude of outrage ('antipathy' in Watson's terms) to one of compassion ('sympathy' in Watson's terms). But this mitigation of initial intuitions fails to achieve purchase on many others. Alternative responses range from a steadfast maintenance of outrage, (for those unmoved by the idea that a horrible childhood translates into a moral free pass even when that child becomes an adult), to perhaps a blend of outrage and compassion in which neither attitude ascends to primacy, and if so, does not remain for long.

Regardless of where in this range our intuitions about Harris fall, his case is markedly different from that of Harold and Lori where we see an enduring

consensus of intuitions about their responsibility status. Furthermore, if we imagine ourselves as immediate witnesses of any of these cases our intuitions about the *agency* of each person, intuitions about their responsibility notwithstanding, then we too find some significant differences in the cases. Harold throwing furniture whilst ducking and screaming colloquial terms for genitalia yields, I think uncontroversially, the intuition that Harold does not currently share reality with other folks in the waiting room. Harold does not have much in the way of custody over his agency. Our intuitions about Lori's agency are similarly uncontroversial. Lori's impairment is immediately obvious, even to the clinically untrained eye. This is, however, not at all true with respect to Harris, Tom and other psychopaths.

As we shall see later in discussing Strawson, the intuition that psychopaths are not responsible for their behavior extends from what Strawson appears to assume is an uncontroversial intuition that the psychopaths lacks agency. If indeed Tom lacks agency, in whole or in part, his lacking is not at all obvious to the casual observer or even members of a parole board. Tom's ability to effectively communicate, including giving reasons for his acting (poor as they may seem to the rest of us), even his ability to appeal to the religious sensitivities of others, not to mention the fact that *he can speak his own name*, belies the notion that Tom's (and other psychopaths) absence of agency is anything remotely like that of either Harold or Lori. Tom's positioning on the responsibility continuum (or 'capacity for morally responsible agency') must then be positioned far closer to full agency than that of Harold or Lori.

A special case must be made then in order to show that psychopaths are completely, or even partially, lacking in moral sensitivity (or in the capacity for moral responsibility). A brief summary of that case runs as follows. Psychopaths, owing to their specific mental illness, lack a very specific and discrete incapacity, namely, that of moral sensitivity. All other capacities remain intact. The onset of psychopathy (whether via developmental trauma or damaged brains, or both) is of no fault by those so inflicted, and so, psychopaths are not at fault for their actions. Psychopaths, it seems, suffer from a moral pathology.

### **Ella the Empath**

If we follow the same line of reasoning with Ella that many philosophers apply to psychopaths (to the conclusion that psychopaths are not responsible or blameworthy owing to their disease of moral insensitivity), then we must conclude that Ella is never praiseworthy for her atypical pattern of generous behavior. Why? Because, as the argument goes, her abnormal pattern of generous behavior is alleged to emanate (spring from?) her alleged disorder, over which she has no responsibility for *having*, or for bringing the disorder about (*she* played no contributing etiological role).<sup>19</sup>

---

<sup>19</sup> Even the preponderance of moral feeling is not be enough to consider Ella's actions as morally worthy in a robustly Kantian sense. Ella cannot rationally will a maxim (perhaps not even a subjective maxim) but can only *merely* act from her sentiments. Yet, however short of moral autonomy Ella may fall, it is another proposition entirely that we pathologize her condition. It is this latter since to which I object.

But it seems odd to suggest that we would be doing something inappropriate in expressing gratitude, or holding any related positive reactive attitudes for that matter, toward Ella when she does something generous. Perhaps this intuitive oddness Ella presents is because philosophers are far less concerned about inappropriate expressions of gratitude, that is, there is little worry about whether expressed gratitude is epistemically and morally justified or appropriate. But if holding persons responsible means to treat persons as moral agents, then we would be holding Ella responsible in expressing gratitude to her inaccurately. Though I can only here speculate, I think this means that the asymmetry with respect to demanding judicious expressions of blame related attitudes must be about something more than a concern for epistemic accuracy. (I am unclear what more this concern is in fact about.) Let's assume that the arising concerns are moral and humanitarian. Among other asymmetries revealed by Ella's case, it shows that we should embrace the general practice of dispensing gratitude far less judiciously than we do resentment, which should be curtailed. But, as we will see, this general practice is not the foundational claim used by philosophers to excuse (and exempt) the psychopath from responsibility. Rather, it is psychopathy the disorder that forms the foundation of the excusing argument. (This is why this dissertation pays close attention to the concept of psychopathy as a crucial component to understanding his responsibility and how we should respond to his wrongdoing.)

Ella's case also shows that our commitment to identifying more and more discrete psychological ailments has limits. Perhaps we should exclude psychopathy from the list of mental illnesses altogether. Regardless, we would never pathologize

generosity or empathy. Furthermore, Ella's case shows that we forego consistency with respect to the demand that we should only hold persons responsible in so far as our responses symmetrically match one's position on the responsibility continuum and that our responses can only be determined *after* one is accordingly positioned on that continuum. We need not form symmetrical responses and we need not, perhaps we cannot first settle (in advance to responding) one's position on the responsibility continuum. Our moral practices, with respect to holding others responsible, are permeated with asymmetry and this need not be viewed as a problem.

### **Conclusion of Cases**

The upshot of case discussion is that while many mental illnesses do yield the intuition that those suffering are not morally responsible for *at least some* of their behaviors, it does not follow (as the pedophile and the psychopath cases show) that all mental illnesses yield the intuition to excuse or exempt the person from responsibility. Furthermore, pedophilia and psychopathy, unlike many other mental illnesses (psychotic disorders, cognitive impairment, dementia, etc.), are not defined by behaviors that are directly harmful to others. So, it would be wrong to draw the general conclusion that mental illness excuse and/or exempt persons from moral responsibility altogether. Alternatively, we might see *some* mental illnesses, more than others, as mitigating the degree of responsibility of its sufferer. Surely there is something desirable about the idea that responsibility comes in degrees and that we should likewise tailor our responses to moral wrongdoing accordingly. But

it does not follow that a person's degree of responsibility be met with a perfectly symmetrical degree of moral response. Sometimes we do issue moral responses that, in effect, demand that the agent extend their degree (or fulfillment) of responsibility beyond one's current performance in meeting basic moral demands.

Mental illness appears to affect our intuitions about agency and responsibility differently than does physical illness. In cases of physical illness we tend to separate the illness from the agent, or at least we do not tend to conflate the person with the illness. But in cases of mental illness we tend to define the *agent herself* with the given mental illness. The agent is no longer 'Mary who has been diagnosed with schizophrenia,' she simply becomes 'Mary the schizophrenic,' or 'Robert the psychopath'. We will see this very view in Strawson as he too takes for granted that mental illnesses replace, or eclipse, a person's proper agency with an ill agency. We also see this view from Levy, as his argument is essentially that psychopaths do not cause themselves in a manner necessary for moral responsibility. Now it does seem plausible to hold that physical illness can change a person's agency. That idea is not controversial. But we don't believe that a person's *agency* is what *has* cancer. Rather, some organ or network of organs *has* the cancer. By contrast, there is no organ or network of organs by which a person *has* psychopathy. Rather, we see psychopathy (along with many other mental illnesses) as enmeshed with, and not distinct from, a person's agency. Herein lies the source of our divided views of the responsibility of psychopaths. If we understand psychopaths as diseased agents, specifically *morally* diseased agents, we tend not to find them morally responsible (since they are not responsible for acquiring the

disease, they are not responsible due to the diseased source of their behaviors). By contrast, if we understand psychopathy as describing a pattern of behavior rather than describing an agent, then there remains room for responsibility. Why? Briefly, because we believe that behaviors can be modified whereas we tend to believe agency to be fixed and static, especially in adults (this is why we tend to view 'self-forming actions' as those occurring earlier in our psychological development *only*, whereas our adult psychologies by contrast are somehow incapable of any further 'self-forming'), and more so, we believe diseased agency to be ultra fixed and extra unchangeable.

I began this chapter by asking: 'what is a psychopath?' and found the answer to be something distinct from what philosophers frequently stipulate. I also argued that psychopathy is a moral concept and that it is identified and diagnosed on specifically behavioral grounds and that this presents good reason to reject the idea that psychopathy can be understood as a medical illness to be explained in empirical, value-free terms. This undermines the notion that psychopathy is a disease that operates independently of its owner *causing* his moral badness. Rather, it is his moral badness that defines his sickness.

## **1.8 Conclusion**

We began with some cases to show how psychopaths, unlike individuals more globally incapacitated by other mental illnesses (e.g., psychotic disorders, severe cognitive impairment), do not invite the pair of intuitions that they lack

agency and because of this they are not responsible for their actions. Psychopaths are cognitively intact, fully oriented to reality and are capable of verbally communicating reasons for a given action and behave with a custody over their intentions that is no different than non-psychopaths, (aside from the intent to harm). Not only is it very difficult for clinicians to identify and diagnose individuals with Antisocial Personality Disorder (especially without a criminal record on hand), it is even more difficult for a random member of society to know whether she is being assaulted by a psychopath or non-psychopath. There is simply no agency and responsibility cancelling intuition with the offending psychopath unlike the cases of Harold and Lori, for instance. Beyond these intuitive observations I provided a provisional argument that the capacity for moral responsibility that Harold (temporarily) and Lori (permanently) lack is a result of a more global lack of capacity. Harold and Lori's lack of orientation is global and their moral orientation washes out with a wider disorientation. This is not so for the psychopath who is alleged to suffer only from the discrete incapacity of moral orientation.

I then moved on to examine the concept of psychopathy within the context of a broader debate over the value-free or value-laden nature of mental illness. I argued that moral values are a core feature of the concept of psychopathy and its diagnostic criteria. Once we acknowledge that psychopathy is defined on morally normative grounds several concerns emerge, including some doubts as to whether psychopathy should be conceived as a disease at all and whether or not pathologizing thoroughgoing immoral behavior represents advancement in our social and moral policies. Comparing Antisocial Personality Disorder to the former

Homosexual Personality Disorder, I argued that pathologizing behavior falling outside conventional moral norms does represent moral progress. As odd or paradoxical as it may seem, the removal of Homosexual Personality Disorder from the lexicon of mental illnesses, thus rendering homosexuality once again fair game for moral assessment, *does* represent moral progress. For, it is better to morally judge homosexuality (one way or the other) than it is to *pathologize* it. The same holds for psychopathy. There are, however, alternative conceptions available in the space between moral judgment and pathologization. Exploring notions of 'holding responsible' that lies within that space is a central theme of this dissertation.

I briefly want to suggest reply to the question, what is a psychopath? Psychopaths are severe and persistent serial wrongdoers most of whom age out of their 'symptoms.' We know that they intentionally wrong people. We do not know why they consistently do so. We simply do not have a causal-etiological explanation for psychopathy. While we might point to various sociological, formative psychological and biological factors involved in the etiology of psychopathy, we need to take care not to presume that we are dealing with a medical disease. We also need to recognize that our moral gaze with respect to psychopaths needs to turn from historical considerations to include the future considerations in considering how we are to respond to current psychopathic wrongdoing. Despite the fact that the behavioral traits of the psychopath remain fairly entrenched deep into his adult life we must acknowledge that his symptoms dissipate. From this observation we can begin to formulate moral responses to the psychopath that are diachronic in nature. This means that we need to re-examine the question of how

should we respond to wrongdoing psychopaths and revise the response that the common view has heretofore instructed.

Some readers may not be ready to sign on to my suggestion that it is a mistake to view psychopathy as a medical disease, nor to the idea that psychopaths possess sufficient agency for moral responsibility. I agree that psychopaths are very undeveloped morally but I do not think that this poor state should dictate our response to wrong doing psychopaths in the manner suggested by the common view. Also, what the cases show is that psychopaths possess far more agency than many other forms of mental illness, including the first two cases of Harold and Lori. This shows that there is not quick route to exemption for the psychopath from our responses to moral wrongdoing. What makes the case of the psychopath so controversial is, in part, that our intuitions divide along two lines: our conceptions of the kind of intending that warrants responses 'with moral content', (and whether or not psychopaths intend in certain ways), and the tacit and accompanying intuition of how we are to respond to psychopaths. There is more to the intuitive picture regarding psychopathic responsibility than the metaphysical intuitions we have about psychopaths but also intuitions (and assumptions) about what it means to hold others responsible. It is this latter concept to which I now turn my focus. If we assume a narrow sense of what it means to *normally* hold others morally responsible, (to warrant affectively charged negative reactive attitudes of which blame and resentment are the paradigm), then I suppose we should not expose the psychopath to such responses. But it is not clear that we should do so as part of our regular *normal* practices at all. What we will find is that sorting out morally

permissible responses in the *abnormal* case of the psychopath reveals options of 'holding responsible' that may well serve as a guide to our *normal* practices which themselves may be in need of revising.

In the next chapter, I turn to the Strawsonian framework of moral responses that include the reactive attitudes and the objective view. After briefly describing Strawson's framework and his comments on psychopaths (and other 'type 2 agents'), I offer some objections to what I consider an overly coarse-grained distinction between the reactive attitudes and the objective view. This analysis will point us to another pair of coarse-grained distinctions. The first of these is the distinction between agents and non-agent that is drawn along the lines of mental health and mental illness. The second of these is the overly coarse distinction between moral responses and 'treatment.' I will argue that there is more practical and moral wisdom to be had by including behavioral and instrumental considerations in our responses to *any* moral wrongdoing, including psychopathic wrongdoing.

## Chapter 2: Strawson's framework

### 2.1 Introduction

We have now examined the concept of psychopathy and reached two provisional conclusions that first, objects to a value-free disease model understanding of psychopathy, and second, that the question of whether or not the psychopath 'is' responsible or not yields a strong intuition in the affirmative. The custody he holds over his cognitively intact agency is without question, especially when compared to some other severe mental illnesses. The burden is thus very high on those arguing that psychopaths are not morally responsible, and furthermore, that we should not *hold them responsible*. So now we turn our attention to examining what we mean by *holding responsible* and what our understanding of this practice means for our responses to psychopaths.

To structure the discussion, I begin with a general summary of Strawson's (1993) framework and how it is alleged to inform the appropriateness of our responses to moral wrongdoing, including to wrongs done by psychopaths.<sup>20</sup> I am not interested here in arguing either for or against Strawson's general theory of moral responsibility but am rather more interested in his framework and in amending an error of his account that yields the wrong verdict on the responsibility of psychopaths. I focus in particular on distinctions drawn by Strawson between agents and non-agents and between the reactive attitudes and the objective view. I then argue that both sets of distinctions are overly coarse and after examining the

---

<sup>20</sup> All references to Strawson in this dissertation refer to his seminal essay (1962/1993) "Freedom and Resentment."

space in between the reactive attitudes and the objective view we find viable options for holding psychopaths responsible. While psychopaths do indeed present a special case that forces us to tailor our responses to his wrongdoing, his case is not so special so as to excuse or exempt him from responsibility.

## **2.2 Strawson's framework**

Strawson contrasts the reactive attitudes (blame, resentment, gratitude, etc.) with the objective view largely defined by a detachment from, lack of, or natural suspension of the reactive attitudes. This distinction between two kinds of psychological response to human behavior is also alleged to track moral responsibility. Strawson himself was of course much more concerned with a larger debate and in showing that the thesis of determinism, should it be rationally and doxastically accepted, would do very little by way of altering our natural reactive attitudes to others with respect to how we are treated by them. I am not interested here in weighing in on this global debate over moral responsibility or assessing the merits of the so-called 'psychological thesis' of moral responsibility (attributed to Hume as well as Strawson).<sup>21</sup> What is more interesting for my purposes is Strawson's reliance on the case of the psychopath – along reference to other mental illnesses – to mark a distinction between agents who are or are not morally responsible. This distinction is alleged by Strawson to be so naturally obvious that, unlike the acceptance of a general thesis of determinism (in which case nobody is

---

<sup>21</sup> Pereboom (2001)

responsible), it can in fact alter our reactive attitudes which in turn excuse or exempt some persons from moral responsibility. Thus, the generally interesting tension between the reactive attitudes and the objective view noted by Strawson is a tension that appears not to occur with respect to the mentally ill.

Within the Strawsonian framework, the more natural and uninhibited our reactive attitudes are the more we are alleged to be dealing with a morally responsible agent, an apt target of our expressed attitudes. (We maintain the intuitive attitudes unless an excuse or exemption applies. Excuses and exemptions are also alleged by Strawson to be, for the most part, intuitively obvious. I discuss excuses and exemptions in more detail below.) Alternatively, the more we find ourselves naturally taking up the objective view the more we are alleged to be dealing with either a diminished agent or a non-agent altogether. Since mental illness, including psychopathy, features so importantly for Strawson in sorting the responsible from the non-responsible, we should note that his 'psychological thesis' relies significantly on a 'psychopathological thesis' of moral responsibility. It is this paradigmatic shift in viewing the psychopath (and other mentally ill folks) as 'mad' rather than 'bad' that is alleged, to represent a development in our ethical thinking. Yet this development is too ungainly to remain unexamined and it is not be a development that we should lightly accept. Why? Because viewing and treating others as an object rather than as a subject exacts as much morally as it is alleged to provide.

In the following section I provide a slightly more detailed sketch of the reactive attitudes and the distinction between the reactive attitudes and the

objective view. To aid in describing the distinction between the reactive and objective attitudes, Strawson relies on the (allegedly) intuitive distinction between agents and non-agents as revealed by the presence of mental illness. (Later in the dissertation I will argue that Strawson's distinctions are too coarse and, like Levy, commit us to an overly narrow sense of 'holding responsible.')

The pair of distinctions between the reactive attitudes and the objective view, and between agents and non-agents works in tandem. That is, they mutually inform each other. I begin with the general structure of the interpersonal reactive attitudes before moving on to discuss how it is that intuitions about agency are alleged by Strawson to bring us to the objective view.

### **2.2.1 Structure of the interpersonal reactive attitudes**

Strawson organizes the interpersonal exchange of the reactive attitudes as follows:

We demand:

1. Of others towards oneself [these are the 'personal reactive attitudes']
2. Of oneself towards others [these are 'self reactive attitudes']
3. Of others towards others [these are the 'vicarious reactive attitudes']

Strawson describes the reactive attitudes as 'inherently human' interpersonal reactions to how we are treated by others. The reactive attitudes are displayed in both actions and attitudes, (e.g., resentment, indignation, gratitude,

guilt, shame, forgiveness, etc.), and arise when we perceive that the basic demand for *interpersonal regard* has been violated. In Strawson's words, "the basic demand is a moral demand; a demand for reasonable regard; a demand addressed to a moral agent; to one who is capable of understanding the demand."<sup>22</sup>

When others fail to meet the basic demand and treat us without regard, we feel and express the *personal* reactive attitudes, say, of resentment and blame.<sup>23</sup> When others fail to meet the basic demand and treat others without regard, as a third party one feels (by way of empathy and sympathy) and expresses (from solidarity) the reactive attitudes *vicariously*. When someone holds a reactive attitude toward me because I have been perceived to violate their basic demand for regard, (perhaps), I feel shame or guilt, and maybe offer an excuse or even seek forgiveness. These are the self-reactive attitudes that are (typically) felt by an offender. To Strawson, the reactive attitudes are a 'commonplace truth' and are of 'central importance' to the lives of humans. "The generalized and non-generalized forms of demand, and the vicarious and personal reactive attitudes which rest upon, and reflect, them are connected not merely logically. They are connected humanly; and

---

<sup>22</sup> A key phrase is contained herein, that one may demand the basic demand ONLY from an agent capable of understanding that demand qua moral demand. I'm not sure that this is right – nor am I sure that Strawson is operating under a narrow sense of 'understand' here. Levy (2007, etc.) seems to think that the knowledge condition for responsibility be very stringent. Later [?] I challenge such a stringent epistemic condition. For the purposes of this section on Strawson I hope to show that a stringent knowledge condition, whilst helping to fore the RA/OV distinction for Strawson, is subordinate to, and not prior, to the 'appropriateness' judgment.

<sup>23</sup> In chapter 4 I take up distinctions between *holding*, *expressing* the reactive attitudes, with the strongest form of expression as contained in bringing sanctions such as punishment. There I engage Levy's call to inhibit our internal 'holding' of the reactive attitudes too stringent (revealing an internal inconsistency in his argument).

not merely with each other.”<sup>24</sup> Such feelings and attitudes arising from interpersonal relationships are part of the fabric of our being and, for Strawson, serve to inform our understanding of moral responsibility.

Strawson also identifies three questions revolving around the reactive attitudes, detailing the third. A brief summary will aid in understanding the shift from holding the reactive attitudes to taking up the objective view, as envisioned by Strawson.

These three questions are:

1. General causes of the reactive attitudes (something like the flavor of the will, the attitudes and intentions, involved in the act)
2. The variations of the reactive attitudes and the *particular conditions in which they do or do not seem natural or reasonable or appropriate*
3. What would it be like, what it is like, to not suffer the reactive attitudes

Strawson is particularly interested in the second and third questions.

### **2.2.2 Commentary on the reactive attitudes**

As stated, question 1 appears to assume the causes of the reactive attitudes to be unidirectional, where the causal onus of the offended person’s reactive attitude rests entirely within the psychological features of the offending person. By

---

<sup>24</sup> Strawson (1993, p. 48), “The central commonplace that I want to insist on is the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about these attitudes and intentions. I can give no simple description of the field of phenomena at the centre of which stands this commonplace truth; for the field is too complex.”

emphasizing the causal role of psychological state of the offending person Strawson may be overlooking the bilateral nature (which is to say, the full interpersonal nature) of the causes of the reactive attitudes and with it, the psychological features of the offended person. Sometimes we overreact even when a person has indeed morally offended us. And sometimes we overreact and insist someone has morally wronged us even when they have not. Overreactions happen on occasions (perhaps something for which we are all guilty of succumbing to at some time or another) but some folks habitually overreact. Suffice it to say that the reactive attitudes we find ourselves taking up are not always the best guide to the moral responsibility of the offending person. Questions 2 and 3 make some progress toward informing 1, but there is more to be said about the causal power involved in 1 as parasitic on the 'offendability' of the person believing she has been wronged (some folks are easily affected and others more 'thick skinned'). For example, I might overreact when a person, deep within a digital coma gazing into her personal device, runs into me, or worse, merges her car into mine. Perhaps we have not yet hashed out the norm for the appropriate response to such behavior. But when we consider the question of the responsibility of psychopaths we are not talking about behavior that resides in the disputed grey area of potential moral offensiveness. Psychopaths violently assault and murder people. I take it as uncontroversial that assault and murder are generally (maybe even always) morally wrong. What I cannot make sense of is the notion that when a psychopath assaults and murders that he has not committed a moral wrong. Nor can I make sense of the view that the frequency and recurrence of his behavior is makes this behavior *less*, and not *more*, morally wrong.

I do not think that Strawson would disagree with these general observations regarding the accuracy of the reactive attitudes. Question 2 clearly addresses the question of the appropriateness of the attitudes. The specific concern I raise here is that the norms we are attempting to establish in question 2, as informed in part by our response to question 1, cannot be fully explained, causally or meaningfully, without consideration of the intrapsychological state of the offended person as it too plays a causal role in the formation, intensity and duration of the reactive attitudes, and by extension, in our understanding of moral responsibility. (I also suspect that in cases of psychopathic wrongdoing and assessing whether or not the reactive attitudes are appropriate is additionally informed by the doxastic attitude of the offended person, *specifically* with respect to the degree of belief in the disease model of psychopathy. If the background assumption in the disease model already has doxastic purchase in the mind of the offended, then the reactive attitudes may not even get out of the intuitive, emotional gates.)

The degree to which we can, or should, exercise control over our reactive attitudes is not obvious. Strawson is not explicit on this question in general terms but he does seem to believe that suspending the reactive attitudes toward psychopaths (and other mentally ill folks) is autonomic, 'at least among the civilized,' as he says. Two questions arise from this observation. First, can we train ourselves to not take up the reactive attitudes? Perhaps we cannot eliminate them altogether but we can learn to mitigate the manner in which these are expressed and maybe even the degree to which we feel them. (I take this as a derivative of Strawson's third question above.) As Strawson notes, even if we could chose to do

so, we might reject living regularly in only the objective view. Second, if we can revise reactive attitudes *out* on occasions when they seem inappropriate, then we might also be able to revise them *in* on occasions when they seem appropriate and we have failed to naturally take them up. By extension, if I naturally take up the objective view towards the psychopath, and that turns out to be the morally wrong thing to do, then I might need to train myself to take up the reactive attitudes toward him. (What we don't yet know is which attitudes we should take up and, more importantly, how or even if, we should express them to the psychopath.)

### **2.3 On the reactive attitude/objective view distinction**

In this section I take up Strawson's distinction between the reactive attitudes and the objective view, including the standards of special considerations by which this distinction is alleged to be drawn. I previously noted that I find the distinction between reactive attitudes and the objective view overly stark as portrayed by Strawson specifically with respect to the role mental illness in making this distinction. To be fair to Strawson it should be noted that he displays sensitivity to the general tension between the reactive and objective views and recognizes that some kinds of agents, like children and 'neurotics,' bring out this tension.<sup>25</sup> Strawson yet remains committed that 'the objective attitudes are appropriate toward neurotic behavior.'<sup>26</sup> The following summary and analysis should be viewed then as equal parts challenge to Strawson's reliance on the mentally ill in attempting

---

<sup>25</sup> Strawson p. 61

<sup>26</sup> Ibid

to distinguish determined from non-determined behavior, and by extension, the appropriate type of attitude we should take up, and in part an addendum to his self-admitted 'speculative' and 'uncertain' thoughts involving the mentally ill, including the 'moral idiot.' (I will be arguing, in part, that Strawson - as well as other proponents of the common view - define moral responsibility in an overly narrow sense as the appropriate purview of the reactive attitudes. My aim then is, in part, to make sense of the idea that we can hold persons responsible without overly relying on the reactive attitudes.)

Strawson claims that the appropriateness of the reactive attitudes is informed by the absence of 'special considerations.' Normal occasions for the reactive attitudes, absent special considerations, are "situations in which one person is offended or injured by the actions of another and in which – in the absence of special considerations – the offended person might naturally or normally be expected to feel resentment."<sup>27</sup> Special considerations thus figure centrally in assessing the appropriateness of the reactive attitudes. Strawson explains that 'special considerations' naturally inhibit (to 'modify or mollify or remove altogether') the reactive attitudes and can be roughly divided into two kinds: knowledge related expressions and volition-related expressions. (Not incidentally, these two conditions correspond to Aristotle's two conditions for moral responsibility.)

1. Knowledge-related considerations: 'He didn't mean to', 'He hadn't realized,'  
'He didn't know'

---

<sup>27</sup> Strawson 1993, p. 50

2. Volition-related consideration: 'He couldn't help it', 'He was pushed', 'He had to do it', 'He had no alternative'

Generic volition-related considerations are sub-divided into situation-related conditions and agency-related conditions where in the former case agency appears to be temporarily suspended whereas it is alleged to be *absent* in the latter.

- a. Situational conditions: 'He wasn't himself recently', 'He has been under great strain recently', 'He was acting under post-hypnotic suggestion'
- b. Agency conditions: 'He's only a child', 'He's a hopeless schizophrenic' [Why 'hopeless?'], 'That's purely compulsive behavior', 'His mind has been systematically perverted'<sup>28</sup>

Special considerations of the kind in 2b are of particular interest to our discussion as the phrases cited here by Strawson figure centrally in the common view, where the presence of mental illness is routinely treated as automatic intuitive grounds for dismissal from moral responsibility; three of his four examples in 2b pertain to mental illness, and he lists more specific disorders, the most important of which for our purposes is psychopath, 'someone wholly lacking, as we say, in moral sense'.<sup>29</sup> On Strawson's view it is patently inappropriate to hold mentally ill agents

---

<sup>28</sup> Strawson 1993, p. 50-1

<sup>29</sup> Strawson 1993, p. 58

responsible. That is, it is inappropriate to hold the reactive attitudes toward 2b type agents *if* the attitudes have not already been intuitively disinhibited. Cases involved in 2b, according to Strawson, ‘invite us to suspend our ordinary reactive attitudes *towards the agent, either at the time of his action or for all time,*’ [emphasis added].<sup>30</sup> However, as we noted in chapter 1, psychopathy is not plainly identifiable as say, Harold in the throes of hallucinations or Lori’s profound mental retardation. Psychopathic behavior does not invite the suspension of the reactive attitudes, as Strawson seems to believe given his inclusion of the ‘moral idiot’ in his examples of 2b type agents. While the presence of moral idiocy may be plainly present in the world what is not obvious is when morally idiotic behavior is committed by ‘normal’ people as opposed to being done by psychopaths. Nor is it clear when we should excuse morally idiotic behavior and when we should exempt those who appear fully committed to moral idiocy. Psychopaths, if they indeed exist, might not be the only humans committed to moral idiocy; moral idiocy is perhaps a far more widespread than we care to admit.

A description of the Strawsonian account is still incomplete. We can complete it by connecting Strawson’s two kinds of ‘special considerations’ with Gary Watson’s (1993) two types of corresponding excuses.

Watson distinguishes type 1 pleas (excuses) from type 2 pleas (exemptions). Type 1 pleas invite us to view the *injury* as something for which the agent was not fully responsible. In Strawson’s words, (the generic volition-related expression is the Strawsonian corollary to Watson’s type 1 plea), such injuries ‘do not invite us to

---

<sup>30</sup> Ibid

view the agent as other than a fully responsible agent,' but merely invite us to view the agent's injurious action as unintentional (or not intentional in the way required for moral responsibility). Special considerations of the generic volitional variety, or type 1 pleas, attach to actions and injuries and do not invite us to question the agency of the offender (or the accidental offender).

Type 2 pleas by contrast, (cf. special consideration 2b), attach to *agents* where agency is called into doubt; type 2 pleas in this sense would rather attach to the absence of agency. In Strawson's words (again), occasions of 2b, 'invite us to suspend our ordinary reactive attitudes *towards the agent, either at the time of his action or for all time*. It is at this point when the impulse of taking up the objective view enters the scene for Strawson. But there is psychological (as well as conceptual and logical) space between (actively or passively) *suspending or inhibiting* our reactive attitudes and then either actively or passively taking up the objective view. For example, on a given occasion, the reactive attitudes may be so disinhibited within me so as to never impinge on my awareness. This inhibition is more thoroughgoing in its absence of the reactive attitudes than if I initially have the reaction which I then suspend. In the first sense I never really had the reactive attitude but in the second sense I had the attitude and then, actively, put it on hold – emotionally and psychologically. (Maybe there is a deeper problem of moral *qualia*, even agency *qualia*, that is fueling disagreement here.)

Strawson portrays the psychological transition involved in these special considerations as rather seamless and intuitive, (though elsewhere, as noted, he tacitly recognizes the tension in us regarding this transition).

“They do not invite us to see the agent’s action in a way consistent with the full retention or ordinary inter-personal attitudes and merely inconsistent with one particular attitude. They invite us to *view the agent himself in a different light* from the light in which we should normally view one who has acted as he has acted.”<sup>31</sup>

In other words, such special considerations are alleged to naturally, intuitively and perhaps uniformly (among all potential reactors) yield the verdict that no normal person would act in such a manner.<sup>32</sup> And when a person’s psychological normalcy is called into question, we detach from the reactive attitudes and adopting the objective view from whence, ‘individuals are not resented or esteemed, but are treated as one to be controlled managed, manipulated, etc.’. For Strawson, members of a civilized society respond to such ‘agents’ by detaching from the reactive attitudes. But what is not clear in this picture, given the alleged intuitive invitation to suspend our reactive attitudes, is whether we as reactors *can* revise this ‘invitation’ to do so. Nor is it obviously appropriate to adopt a disposition toward the psychologically ill as someone only to be ‘controlled, managed, and manipulated.’ In fact, one might experience a Strawsonian

---

<sup>31</sup> Strawson 1993, p. 58

<sup>32</sup> There may be a distinction between the ‘Not a normal person!’ intuition and the ‘Is mentally sick!’ intuition. Which of these two intuitions one might have when being assaulted by an individual may be distinct on a given occasion – and much of such intuitions will be filled by context and the *seeming* connection of the offending individual to reality.

involuntary reactive attitude of disgust at taking the very disposition toward the mentally ill that is entailed by taking up the objective view. Why? Briefly, the objective dispositions as so described are deeply disrespectful to persons and it is not obvious that the mentally ill lack personhood or agency in the manner portrayed by the Strawsonian tradition. The ready common view reply here of course is that respect is reserved for persons and the individuals as described in 2b are neither agents nor persons in the full sense, and so, where there is no 'person' there is no disrespect when manipulating. Or, if you prefer, respect is reserved in the degree to which agency is fulfilled (or not). But this *moral* inference does not follow, as the relationship between agency and moral respect is neither straightforward nor symmetrical. Those who fall short of full agency, even those mostly lacking agency, should nonetheless be respected morally. So, the common view suggestion that we best respect non-agents by treating them as non-agents is not the right moral response.

Furthermore, as a dialectical objection, 2b appears to beg the agency question out of the scenario from the outset. The absence of agency is not obvious with psychopaths, which suggests that a natural inhibition of the reactive attitudes may not be signaled to us in a way that occurs in other type 2 cases (e.g., as with Harold or Lori). In particular, it is the instrumental aggression exhibited by the psychopath that undermines the suggestion that his agency is obviously absent. The psychopaths' capacity to manipulate others belies the suggestion that he lacks agency.

What we have before us now are questions relating to the relationship between agency, mental illness and what the objective view licenses in terms of our interactions with people represented in 2b cases. It is also not clear that the reactive attitudes, as Strawson suggests, are less appropriate than taking up the potentially disrespectful objective view toward the psychopath. It is also not clear that the reactive attitudes themselves do not carry, at least tacitly, an instrumental goal to manipulate or modify behavior. I shall argue that there is a deep, almost inescapable instrumental component involved in the reactive attitudes and this exposes a weakness in one manner alleged to distinguish the reactive attitudes from the objective view. The tension of which type of attitude we should adopt in 2b type cases reveal questions about the general propriety of both the reactive attitudes and the objective view that, when addressed, will in turn help to resolve the RA/OV tension. Nonetheless, since much hinges on the role of mental illness as agency undermining, and thus, informing the appropriateness the type of attitude we are to take toward up must be examined cautiously so as not to beg the question of agency as involved in 2b cases.<sup>33</sup>

---

<sup>33</sup> In fact, I worry that once we engage the question begging of agency problem that we can't resolve it – and so, we are forced to adopt an *as if* approach to cases involving 2b, accept that the question has been begged ('agency in') and show that this is superior to the alternative, 'agency-out' question begging. The question-begging cycle has to be broken somehow, and owing it is one way – though this is entirely unfashionable for philosophers.

## 2.4 Concerns with Strawson's account

Two sets of concerns with the Strawsonian view emerge from the preceding discussion. The first set pertains to questions regarding *how* we are to identify cases in which it is 'appropriate' to inhibit our reactive attitudes and instead take up the objective view. This task is complicated by the requirement that we correctly, and rather immediately, identify the types of 2b agents listed by Strawson, including psychopaths. How well we can be expected to identify them and how voluntary our ensuing attitudes, or inhibition of them, is not as immediately apparent as Strawson presents. (Antisocial Personality Disorder, along with all of the so-called Personality Disorders, is notoriously difficult for clinicians to catch in assessment and evaluation. Simply put, we are rarely aware of the presence of a psychopath even when we are in the same room with one.) Psychopathic behavior in particular is not the kind that invites a simple inhibition of the reactive attitudes; in fact just the opposite occurs, and for many of us, *even when we learn the offender is a psychopath*.<sup>34</sup> The second set of questions relates more directly to the general appropriateness of the reactive attitudes and of the objective view, aside from special consideration of the psychopath. The alleged special connection between the reactive attitudes and moral responsibility that is assumed to not to hold

---

<sup>34</sup> The 'how' question arises in response to the Strawsonian assumption that the inhibition of a reactive attitude is intuitive, is immediately phenomenally impinging (my term). But 2b cases do not produce such ready-made inhibitions. Rather, it is more likely that our intuitions are informed by a tacit model (in this case, a model of psychopathology as automatically entailing involuntariness), by a theory dependent intuition, if you will. Just as we are inculcated to hold and express certain reactive attitudes, we are inculcated by theoretical paradigms – a few have a stronger grip on our consciousness than does the disease model of mental illness (viz., the very idea that a person can have, let alone be born with, a diseased personality).

between the objective attitudes and responsibility is a bifurcation that we need to overcome. The reactive attitudes have cornered the market of moral responsibility for too long, and the purpose (even the value) of the reactive attitudes is not beyond evaluation and revision. We might also ask related questions regarding the value of the objective view. What is it that makes the objective view a morally superior option in, for instance, 2b type cases? It might not be. What we often find is that the objective view is a resource that can be put, even unconsciously, to pernicious use. Ideally, what we want are responses to psychopathic wrongdoing that make use of the best resources that both the reactive attitudes and objective view have to offer.

#### **2.4.1 Agency recognition and the tension between the reactive attitudes and the objective view**

How are the offended to recognize when the reactive attitudes are appropriate or, alternatively, when the objective view is appropriate? As Strawson describes, special considerations initiate this shift in us from the default position of the reactive attitudes to the objective view, and our recognition of agency is central in steering this shift. This phenomenological shift that moves us from the reactive attitudes to the objective view appears to be twofold. First, the recognition of agency, or more importantly, recognizing the absence of agency, is alleged by Strawson to just be apparent. But the absence of agency is not so apparent in the case of the psychopath, especially not on first glance (and maybe not on second glance either). Second, the recognition of agency is then supposed to steer our reactions and attitudes accordingly. Now, since Strawson holds firmly that the

reactive attitudes neither permit nor require rational justification, (perhaps because they are inherently *emotional* - a term Strawson seems loathe to use), and the presence or absence of agency is supposed to just be apparent to us, then it is very unclear how voluntary or conscious our attitudes towards others actually are. So, the appropriate formation of our attitudes relies heavily on our getting the agency assessment correct. And since our interpretation of behavior plays a crucial role in our often quick, if not hasty, assessments of agency, it is very difficult to get both agency assessments and the appropriate set of attitudes right. (For the time being I set the agency and behavior relationship to the side but will return to argue that our reactive attitudes might be better directed by separating two distinct *reactans* in behavior on the one hand and agency on the other.) Given these difficulties, the amount or degree of active control one can have over suspending the reactive attitudes and the related degree of active control one can have in taking up the objective view is unclear. This seems, on pain of parity, to require that ensuing assessments of the appropriateness of the attitudes actually taken up also take the interpretive and intrapsychological complexities that inform our attitudes into account.

This highlights a false structural assumption that psychiatric ethics is done much like medical ethics in which we assume we are first, dealing with clearly descriptive, factual, medical domain in psychiatric concepts as we are with non-psychiatric illness *to which we then formulate normative responses*. But this analogous ethical approach is false. It is false because, as we see in chapter 1, when it comes to psychopathy (and psychiatric concepts in general) we are not dealing

with a descriptive-medical concept, we are dealing with a morally normative concept *to which we then must formulate normative responses.*

Since Strawson is working from a standard of attitudes taken up in 'normal' relationships between two 'normal' adults that he argues do not to apply in cases in which one agent is abnormal, then it is important to take a moment to discuss this standard. After which we can then assess the general value of this standard as well as its import to responding to psychopathic wrongdoing. We may find that our responses to psychopaths, contrary to the common view, need to be more like normal responses and include the reactive attitudes. Alternatively, our heretofore-normal responses to each other should (and do!) include more objective attitudes. Perhaps the abnormal standards and the objective view should be applied more generally in our moral practices.

#### **2.4.2 On the general propriety of reactive attitudes**

In this section I question what is alleged to be special about the content of the reactive attitudes specifically with respect to why they are supposed to be reserved for only a certain class of agents. (This class of fully responsible agents, for whom the reactive attitudes are to be reserved, may be much smaller than we care to admit.) I also examine the degree to which the reactive attitudes are instrumental and not merely a reflection of our inherent human nature. If the reactive attitudes (irreducibly) contain an instrumental function, as I argue they do, then this instrumental feature of this set of attitudes shares a key feature with the objective

view, thus rendering the distinction less clear than portrayed by Strawson. (This also requires clarification of the ‘detachment’ that is alleged to occur in moving from the reactive attitudes to the objective view. I take up the concept of detachment and its involvement in the attitude distinction briefly below and more directly in chapter 4.)

On the Strawsonian tradition, the reactive attitudes are appropriately reserved for full members of the moral community, (provided an offending member does not have a type 1 excuse available).<sup>35</sup> Put another way, the reactive attitudes are reserved for agents whom we have *not* ‘come to see in a different light.’ Recall that, on Strawson’s account, we come to see psychopathological agents ‘in a different light,’ whom we ‘set apart from normal participant reactive attitudes,’ and toward whom it is ‘civilized’ to take up the objective view. Though we will visit Strawson’s conception of agency in 2b cases in more detail below, for the time being we can roughly distinguish appropriate targets of the reactive attitudes as ‘normal’ individuals, for whom none of the special considerations apply. One must hold some agential privilege to be appropriately subjected to the reactive attitudes and the agential status is alleged to precede, and thus steer, the formation of our attitudes. Suppose then that we momentarily set aside special considerations concerning agency and imagine an appropriate interpersonal exchange involving the reactive

---

<sup>35</sup> Strawson 1993, p. 58. Such ‘free’ agents are described by Strawson as “nothing but the absence of certain conditions the presence of which would make moral condemnation or punishment inappropriate.... ..conditions like compulsion by another, or innate incapacity, or insanity, or less extreme forms of psychological disorder, or the existence of circumstances in which the making of any other choice would be morally inadmissible or would be too much to expect of any man.”

attitudes. What is it that these inescapably human, non-detached interpersonal attitudes *do*? Put another way, one might ask, ‘what is the purpose of the reactive attitudes?’ (I take this question to be distinct from the three questions raised by Strawson, outlined above, section 2.2.1. In taking up the question of the purpose of the reactive attitudes, one might additionally make some headway on the first and second questions.)

To Strawson’s ears the very question, ‘what is the purpose of the reactive attitudes’ may seem question-begging and misguided from the outset, as the reactive attitudes are alleged to be inherently human such that they are rendered unrevisable.<sup>36</sup> Question begging notwithstanding, then, it does seem fair to raise the question whether or not the reactive attitudes contain an instrumental component because this component is alleged to partly inform the distinction between the inherent nature of the reactive attitudes as contrast with the social utility of the objective view. If there is such a component to the reactive attitudes then one manner of distinguishing the contrasting sets of attitudes is lost and, perhaps, along with it, how we divide responsible from non-responsible agents.

In what ways then might the reactive attitudes include in them some instrumental purpose? Perhaps the reactive attitudes have served an evolutionary purpose for humans. Levy (2010) provides a compelling bio-psychological

---

<sup>36</sup> Strawson; the passage straddling pages 60-1 is particularly relevant to this discussion where Strawson entrenches in the contrast of attitudes by strictly limiting the reactive attitudes to a ‘pure and unqualified form.’ What this means, *operationally*, in describing the reactive attitudes is left so under-described by Strawson that it comes across as a purely rhetorical phrase.

evolutionary account of the function of the reactive attitudes.<sup>37</sup> Perhaps the reactive attitudes (in all forms, personal, self and vicarious) have aided humans as a species. Perhaps the execution of any of the individual forms of the reactive attitudes present an adaptive advantage to individual members of our species. Perhaps human reactive attitudes are similar to the utterances or cries of other non-human animals. On the one hand, Marmots seem particularly adept at raising alarm for the betterment of the Marmot community but on the other hand, grieving Pilot Whales seem to be at an evolutionary disadvantage, as they tend to beach and die. If there is such an evolutionary explanatory account of the function of the human reactive attitudes, it is not clear what bearing this has on our assessment of the normative value of the reactive attitudes. We may be hardwired to have reactive attitudes but we are also hardwired to eat sweet, salty and fat foods yet we revise our nutritional intake, at least those of us fortunate enough to afford to do so. So, whatever evolutionary account there is to be given here, it does not seem to dictate our ability to revise something about the practice of *expressing* the reactive attitudes. Whether or not we can revise *feeling* or *holding* these attitudes is a separate matter. There may be a boundary of voluntariness that separates our ability to revise how we express the reactive attitudes as distinct from our ability to hold them internally.

Recall, however, that in Strawson's picture, the reactive attitudes are a

---

<sup>37</sup> One might wonder what, if any, adaptive advantage of psychopaths, and *a fortiori*, an absence of the self-reactive attitudes. IF we take psychopaths to be a natural kind (*taxon*) – though there are good reasons to reject this view – we might take their continued existence to reveal that the self-reactive attitudes aren't necessary for propagation. Of course, it is not clear what it is that an evolutionary account of our moral life means for the normative practices of the reactive attitudes. Here I think it is safe to say that Strawson would reply, 'none.' On this issue, I agree.

commonplace truth, are inherent in us and are robust enough to carry on in our interpersonal practices, even in the face of a (potentially) true thesis of determinism.<sup>38</sup> While these observations may well provide empirical evidence that bolsters Strawson's claim that the reactive attitudes are inherently human, this is not precisely the sense of the instrumental component of the reactive attitudes, which I suspect obtain.

There seems to be at least two additional senses in which there is instrumental purpose to the expression of the reactive attitudes (for now, the 'self reactive attitudes'). The first of these is the reactive attitudes as catharsis, for the purpose of the offended to emote, or discharge intrapsychological tension that accompanies one's sense of being offended. The second is that the (self) reactive attitudes are a demand for behavioral and attitude modification on the part of the offender (we might alternatively synonymously call this an educative function, but my emphasis on behavior will become increasingly apparent). Put another way, the self-reactive attitudes demand that the ill will revealed by a wrongdoer in her wrongdoing be revised. Apologies and forgiveness, if genuinely willed and expressed, put on display a will that is, or appears, sensitive once again, to the basic

---

<sup>38</sup> What this means for the rationality of the reactive attitudes isn't clear, but does not appear to bode well. If the reactive attitudes, and our alleged *necessity* for them, are impervious to reasons (say, as might be offered in propositional form in the thesis of determinism), then ought we as hyper-rational philosophers be categorically skeptical about the value of the reactive attitudes? Besides, is not the history of philosophy replete with reinforcement that we *be* more rational than we frequently are? It is. Nonetheless, there seems to remain a lingering concern about the voluntary nature of the reactive attitudes and whether we can revise them in their variation and expression, or more, whether we can (let alone whether we would choose to if we could) revise them out of our practices altogether. Paraphrasing Augustine: 'Lord, delivery me from resentment... only not just yet.'

demand for interpersonal regard. At the very least, reactive attitudes seek *in the offending party* behavioral change and only ideally, a 'deeper' corresponding attitude modification to match. I now take up these two instrumental purposes of the reactive attitudes in more detail.

The emotive content contained within a reactive attitude, and the purpose of its expression might simply be a cathartic discharge of psychological tension. The instrumental purpose here is only about the offended party. But, reactive attitudes are impulses that can endure as, frequently and unfortunately, seems to occur particularly so with the paradigmatic attitude of resentment.<sup>39</sup> Enduring resentment is poisonous, however, and it may be practically rational for the offended, when offended, to cathart and without delay so as to rid the toxin. (If only this actually worked for this purpose.) The psychological release involved in the reactive attitudes may justify the offended party in expressing her reactive attitudes simply for this reason even if it harms the offending party. But this may not be justified. If the upshot of expressed resentment is to exact an emotional cost and to achieve (or restore) an interpersonal emotional *balance* it may not obviously be appropriate, particularly if this emerges from an attitude of revenge. It is not true however that we must, or even should, requite a moral offense with emotional harm. Buddha, Jesus and Nietzsche all speak against this tendency. So, whilst Strawson finds it generally acceptable (absent special considerations) to express the

---

<sup>39</sup> Nietzsche has a lot to say about resentment and revenge, most of which is to lament its role in the history of human values. Resentment in his eyes occurs only among the weak spirited with low tolerance for psychological tension coupled with a strong impulse to punish.

reactive attitudes to others when failing to demonstrate the basic demand for interpersonal regard, it does not simply follow that reciprocating emotional harm (by blaming and resenting) is appropriate, even among the mentally healthy. An expression of blame or resentment may itself occur *without regard for the basic demand*. And if such practices are to be reserved only for meriting members of the 'moral community' then one might rationally decline membership in a community defined, in part, by the impulse to maintain an interpersonal emotional account that *must balance*.<sup>40</sup> For, this cathartic component of the reactive attitudes holds, almost exclusively, the interests of the offended party.<sup>41</sup>

---

<sup>40</sup> Additionally, many of the negative reactive attitudes share a similar interpersonal transactional structure with revenge. In such cases the owner of psychological tension (revenge) find the tension unbearable and seeks release – and only onto another will, another person. For the vengeful, it's not enough to rid one's psyche of the tension, moreover the tension must be transferred to another psyche, and preferable to the one viewed as the initial offender. Such is the type case with revenge. But we might replace 'revenge' with 'gratitude' for instance to see if a similar intrapsychological-to-interpsychological structure would occur. But here a sincere warning is in order. Notice that I've spoken about 'type' attitudes (revenge, resentment, gratitude) as if these phenomena are patterned across individuals and that we would all experience these type attitudes equally, whether in intensity or in the need to release the attitude. But this is wrong. So, here we might rephrase Strawson's second question above (I.b) and not ask about the variation of the reactive attitudes themselves but about the variation of *reactors* and how the type attitudes are experienced *unequally*. We do not all participate in the same number of temperaments, nor with the same intensity. And so, we cannot discuss the 'human reactive attitudes' (as Strawson does, and as I have) as if they are to be experienced (*qua* type) across people equally and monolithically. This is what makes generalizing about the 'appropriateness' (Strawson question 2) of a reactive attitude type on a given token occasion. And this same generalization problem holds for the query into the 'causes' (Strawson question 1) of the reactive attitudes. Yet another reference to Nietzsche seems appropriate here, sp. *On the Genealogy of Morals* where the origin and development of two distinct moralities, of masters and of slaves, is charted. In fact, one might just revise the title of the book to read *On the Genealogy of Moral Psychology*, as the fundamental distinction drawn is between the strong and weak willed (or strong and weak psychologies) and the moral systems

The important point here, however, is not to evaluate the appropriateness of the cathartic nature of the reactive attitudes but rather to acknowledge its involvement in these attitudes, and more importantly, to acknowledge the instrumental nature involved in this catharsis.

Beyond the cathartic discharge of psychological tension, the reactive attitudes also serve as reinforcement of moral demands and this frequently entails a demand for change in behavior and attitude. Ideally, expressing the reactive

---

that emerged from those *as a result* of the relative psychological strength or weakness. In his eyes, Christianity arises from slave morality to become the 'shining jewel of resentment', and a 'morality of the hangman.' And the 'moral community', which is envisioned as secular as it appears in current philosophical scholarship, seems too similar to its Christian predecessor for comfort to some of us. And so, with Strawson's third question in mind (what it would be like not to *suffer* the reactive attitudes), the hidden premise of his solution is whether or not we can be *utterly rational* (how Socratic) in order to *not suffer* the psychological tension of emotion that imbues the reactive attitudes. But there is another way to address the third question, not with a hyper rational attempt to not suffer, (Socrates' dying words are proof that this is NOT the approach to the good life, and he himself realized this) and rather with intrapsychological strength that does not share either the impulse to extreme rationality nor to reject suffering as immediately as possible. From Buddha and Nietzsche we learn to 'not requite enmity with enmity' and rather 'show our enemy that he did us some good.'

To connect this footnote more directly to the forthcoming discussion of the responsibility of psychopaths, contrary to Levy's generalized claim that it is 'wrong to blame the psychopath' for his harmful violence, we can't determine propriety across reactors. Victims of psychopaths have no onus to refrain from blaming their assailants. 'We' hyper rational academics removed from the consequence of psychopathic violence *might* be generally inappropriate in blaming psychopaths (though this is not obvious to me whatsoever). But we might have a duty to express vicarious reactive attitudes out of solidarity with his victims.

<sup>41</sup> The reactive attitudes *qua* psychological tension may be bad for the 'sufferer.' Example: prolonged resentment is like a fly in the bottle of our psyche. It sours persons. Early catharsis as opposed to delayed catharsis may be preferable. Non-prolonged catharsis may have the added value of being prone to voluntary release as delay often results in repressed tension that is in turn sublimated and released randomly and unconsciously. Perhaps we should seek a noble 'moral forgetting' as members of the reactive moral community. Some should, but we can't demand that all should.

attitudes not only avoids adding moral misery to the world it results in both behavioral and attitude changes of the offending party. I say ideally because we prefer for the latter to consummate the former. But often it is the case that a moral remonstrance does not yield change in the offending party, neither in his behavior nor his attitude. This lack of response by the wrongdoer is something shared by psychopaths and non-psychopaths alike. And even when things go well, the gains are limited and yield only behavioral changes, that, only if rehearsed will, in Strawson's own words, 'modulate toward true performances.' (Plato, in *Laws IX*, advocates this exact approach in attempting to rehabilitate the incorrigibles – by mimicking the virtuous person until the practice is more genuinely possessed.) The more weight we put on special interpersonal relationships the more we want others to fully internalize behavioral modification with a matching attitude. We want our offenders to *feel* differently, and feel differently *toward* us (we want a *change of will* to occur in our offenders). But complete modification (external behavior and internal attitude) may be too much to ask of some agents. Adolescents and psychopaths are two examples. In these cases we are happy to accept behavioral modification regardless of the 'truth' of the behavioral performances as verified by matching attitude.<sup>42</sup> If expression of the reactive attitudes has the intentional

---

<sup>42</sup> The well-rehearsed, but still ill advised, failed tactic of parents of teenagers behaving badly is to *demand* an attitude change (to demand a certain intrapsychological constellation) *within* their adolescent. Parents often want to break the defiant will of their teenage children, and usually with the interest of moral improvement in mind. But strong willed adolescents will go to war against such an approach. On some occasions a behavioral modification absent a matching attitude change *must suffice*. Practical wisdom dictates that we tailor the demands for modification and this means we proceed with attuned reactive attitudes, and

instrumental purpose of behavioral modification, then we might gauge the appropriateness accordingly.

When the reactive attitudes are expressed with the purpose of seeking behavioral and attitude modification of the offender they take on a similarity to the objective view as a kind of manipulation and control of the offending party. We must be careful with our terms here though as ‘modification’ sounds tame in comparison to the nefarious ‘manipulation.’ But why naturally associate the nefarious ‘manipulation’ with the objective view whilst associating mere ‘modification’ with the reactive attitudes? (This question turns for Strawson on *agency*.) As we just observed, there is often much emotional venom contained in a negative reactive attitude. It is emotional venom we want removed from our psyche to be put in another psyche. It is, in a sense, to treat the offending party as an emotional gutter, in which we often want those emotions to pool as well. If you want to call this ‘modification’, that’s fine. But when so expressing the reactive attitudes we are not seeking that the offender simply cease the offending actions, we want this no doubt, but we also want more. We want an intrapsychological change to occur within the offender (and often by means of harm).<sup>43</sup> Hence the cries, “I

---

without abandoning them altogether in retreat to the objective view, as this is alienating.

<sup>43</sup> ‘Suffering and sanctions’ are the terms used by Fischer and Tognazzini (2012) for instance. Here too the authors seem primarily concerned with caching out ‘moral responsibility’ for the purposes of ‘knowing’ when punishment is justified. The inability to emotionally affect, even harm, psychopaths is what many find so puzzling, and concludes that psychopaths have ‘no moral sense.’ But this view can only occur against a desert-based morality in which ‘normal agents’ are busy seeking to balance the emotional and moral ledger. ‘We can’t exact a moral-emotional cost from psychopaths so they must not be like us.’ But this is a *post hoc*

want him to suffer like I've suffered. I want him to know what it *feels* like." Of course, when we are talking about psychopaths, given their sangfroid callousness, this exchange yields nothing. What both fascinates and frustrates us about the psychopath is that they do not respond to this ritual in the way the rest of us do. (This is of course true for many non-psychopaths as well, e.g., sexists, racists, corporations.) In fact, psychopaths are so skilled at remaining impervious to emotional venom, even when expressed in moral language, that the common view infers from this lack of reciprocity that psychopaths are simply incapable. The inferences mount quickly, and incorrectly, that this incapacity must be physiologically based and further, paradoxically, that psychopaths can be harmed when we blame them. Both claims cannot be correct. (I take up both arguments in chapter 3.)

Two distinct objections to these observations about the instrumental nature of the reactive attitudes lurk. One is a Strawson inspired objection and the other is a Levy inspired objection. I raise both here briefly only to note them; chapter 4 contains a closer discussion. Strawson will likely object that portraying the reactive attitudes as means seeking by emphasizing the instrumental function of them requires ignoring their inherent nature (that in part makes them impervious to the thesis of determinism even if it is true) and to do this is to commit a kind of consequence mistake. My brief reply is that we should no longer ignore two

---

argument. It might not be the case the psychopaths are pathological at all and rather than they have rationally transcended the emotional warfare endemic to our moral community. The problem with psychopaths is that they assault and kill other persons and only tacitly should we be concerned with their alleged physiological disability to have their moral attitudes modified.

important ways in which the reactive attitudes contain in them, perhaps even inherently, the impulse to have an effect – on ourselves and on those who offend us. The real challenge then is the need for practical wisdom in tracking the effects our various attitudes do indeed have. (In chapter 4 I show how this can be done by more carefully distinguishing our sets of *reactans* and *reactanda*.) Catharsis, owing to its chaotic and unwieldy nature, functions much like cluster bombing; it has an effect on others for sure, but the damage is not suitably contained – not in a morally preferred manner anyway. We *should* want our reactive attitudes to target and initiate change and to be *more* not less effective as well as morally sensible. If we give up this instrumental function of the reactive attitudes, then the expression of the attitudes is only about easing the psychological tension of the offended party, and the very act of moral judgment in this scenario is itself merely an exercise of self-indulgence.

The second Levy inspired objection is that we must first *know* whether or not an agent is or is not morally responsible before engaging in the reactive attitudes and not treat them merely *as if* they were a morally responsible agent. In reply, the cases discussed in chapter 1 show that the psychopath does not present in us any doubt that his agency is intact. If the psychopath lacks agency it is certainly not apparent – and often not even to skilled clinicians. We might consider a more charitable approach, however, and agree that the agency question with respect to the psychopath remains open. Once in this position, we must form a response to psychopathic wrongdoing based on either an ‘as if’ or ‘as if not’ assumption. (I argue in chapters 3 and 4 that the ‘as if’ approach is the practically and morally superior

option.) Besides, it is practically unachievable that we first establish agency (an intractable conceptual problem) *prior to* holding and expressing our reactive attitudes.

The upshot of the preceding discussion is that the propriety of the reactive attitudes even between two healthy agents is a standard of holding others responsible that is not without its problems, even when propriety does not turn on the intractable problems of assessing agency. The negative reactive attitudes might be inappropriate regardless of whether special considerations apply. The reactive attitudes seek to bend the will of another and sometimes this takes on a demand for modification and at other times a manipulative character (especially when we want to put someone to shame). We also see that the reactive attitudes contain an instrumental function and, once acknowledged, one component alleged to divide them from the objective view has been shown to rather *connect* these sets of attitudes. Here we might rest with noting that the alleged goods of the reactive attitudes are not beyond criticism and that often when expressed, the instrumental purpose of the reactive attitudes appears to seek similar ends sought in the objective view: to control and manipulate others. When doing so, we never engage other *subjects* but only other *objects*.

### **2.4.3 On the general propriety of the objective view**

This section has three aims. The first is to summarize Strawson's account of the objective view, including its alleged value and the manner in which we arrive at

this view (much of which has been discussed and doubts have been raised).

Following this summary I raise some objections that are designed to show the limits of the value of the objective view as portrayed by Strawson. When we consider the potential *impropriety* of the objective view we can better appreciate the downside to this resource and see that, just as we can wrongly hold the reactive attitudes, we can likewise wrongly hold the objective view. More specifically, we can wrongly refrain from the reactive attitudes (and treating the psychopath as the common view commands shows this mistake). Finally, this section concludes with examining the nature of 'detachment' as involved in taking up the objective view. Strawson is not clear *from* what it is that we as reactors are supposed to detach. I consider some options and suggest a shift in focus from the concept of 'detachment' to the role of affect in both sets of attitudes, the reactive and objective. Herein lies another way in which these sets of attitudes are connected.

To Strawson, the suspension of the reactive attitudes and the ensuing move to take up the '*profoundly* opposed' objective view toward the abnormal is a *consequence* of our seeing the agent in an abnormal light.<sup>44</sup> By now, however, we see that in the case of the psychopath there is not an obvious absence of agency that nudges the attitude shift. (Our intuitions still divide with respect to specifically moral agency.) This invitation to inhibit the reactive attitudes is not prescient with psychopaths as it is, for example, in the cases of Harold and Lori discussed in chapter 1. Suppose, however, that we follow Strawson's intuitions for the moment. When a person's psychological normalcy is called into question, we detach from the

---

<sup>44</sup> Strawson p. 52, 58

reactive attitudes and adopting the objective view from whence, ‘individuals are not resented or esteemed, but are treated as one to be controlled, managed, manipulated, etc.’<sup>45</sup> Using the ‘moral idiot’ as his paradigm case, Strawson insists that resentment toward him is naturally inhibited, which “tends to promote instead the purely objective view of the [abnormal] agent as one posing problems simply of an intellectual understanding, management, treatment and control.”<sup>46</sup> Let us be clear about Strawson’s view with respect to psychopaths and that the appropriate view to take toward him is to be a thoroughgoing and *purely* objective view because, “the moral idiot is, as we say, wholly lacking in moral sense... not one toward whom the demands and expectations lie... when we speak of moral obligation; he is not, to that extent, seen as a morally responsible agent.”<sup>47</sup> (Recall that in the introductory chapter that I portrayed the common view, which finds its roots in Strawson, to see the psychopath as completely and incorrigibly morally blind.) I have already raised a series of concerns about how unlikely it is that we see the psychopath as a non-agent, especially not with any immediacy. If that is the case and the psychopath indeed possesses some degree of agency, then taking the objective view toward him is wrong in a perceptual or epistemic sense, but it is, in addition potentially *morally* wrong to view him in as stark terms as Strawson portrays. I turn now to examine the impropriety of the objective view.

One way to see the downside to the objective view is via Strawson’s application of this view in relatively normal relationships ‘as a resource’ available to

---

<sup>45</sup> Strawson p. 59

<sup>46</sup> Ibid

<sup>47</sup> Ibid

'relieve ourselves from the strain of involvement,' where the agency of the other party is not in question.<sup>48</sup> Strawson doesn't provide a clear example of what he has in mind in these cases, but does suggest that the agency of the 'offender' is not clearly in question. Perhaps 'relief from the burden of involvement' that objective view provides goes badly, for example, that occurs between homosexual children of fundamentally religious parents, where resentment quickly devolves into studied indifference; in this case the objective view may indicate a moral failure. Likewise, such refuge from burden may be a gradual slide into a new default position of a couple together too long and rather than parting ways, remain in a kind of suspended animation. However, such relief is more understandable, say, as something that adult children of dying parents must adopt in the face of their own deteriorating narrative and when the demands of care force a suspension of grief.

A moral evaluation of the objective view in each of these examples will not yield the same verdict. The alienation that occurs in the first two cases is a result of taking up the objective view. It is the *source* of the alienation in a way that it is not in the last example. This means that we need to be very careful in the way we both use and commit to the objective view, even when the slide into that view feels involuntary. Since alienation divorces us from relationships (whether wholly or partially) we must take care not to perpetuate alienation in our use of the objective view as it can be used both as a resource and a weapon. Furthermore, if taking up the objective view plays a contributing role in diminishing the nature of a relationship, it might likewise contribute to whether we see the agent as responsible

---

<sup>48</sup> Strawson p. 52

or not. The relationship between alienation, relationships, the objective view, agency and responsibility may well be circular in nature and not linear. The common view appears to assume this linear view and that agency (and its absence) is the starting point that causes the alienation, and then the diminished relationship, thereby prompting of the objective view and the exempting from responsibility. The assumption that all of these factors rather neatly fall into line is false, of course, and this mistake *begins* with many incorrect assumptions that philosophers hold about the nature of mental illness, psychopathy included (hence the importance of the discussion in chapter 1). We should note that the objective view, because of the ‘detachment’ that defines it, can itself be an obstacle in relationships, and thus in our understanding of responsibility. Since the notion of detachment plays such a crucial, if understated, role in the set of attitudes we take up, I turn now to its role and value.

The notion of detachment is the last on the list of pitfalls of the objective view. I first say what I take Strawson to mean in his use of the ‘non detached’ reactive attitudes as opposed to the ‘detached’ objective view. I then highlight some unwanted implications of detachment.

Strawson is not explicit in describing what exactly it is that we are detaching from when we take up the objective view. There are (at least) two components involved in detachment: emotions and agents. Since we have just discussed the issue of detachment and agency, I will primarily focus on the role of emotional detachment in the attitudes. It is in the sense that I take Strawson to be primarily using the notion of detachment. (Mental health practitioners tend to use ‘affect’ in place of ‘emotion.’) When we take up the issue of the emotional magnitude of a

given attitude, which in turn inform our assessment of the appropriateness of the attitude, we must be clear that we are in the domain of the emotional and intrapsychological states of the reactor. Here, Strawson cites the view a therapist takes up toward her patient as a model for describing what occurs when taking up the objective view. But, I think Strawson is quite wrong in his depiction of the therapeutic ideal by explicating it in terms of the objective view, and thus, appears to beg the question that the objective view is the right one to take toward type 2 agents. Let me further explain.

While it is true that mental health practitioners are trained to reign in our affective responses, at least *revealing* our affective responses, we are also trained to use them as a therapeutic tool, including the building of a genuine human relationship with our 'patient.' We therapists are also trained to use our affective responses as a diagnostic tool. Perhaps most importantly, we therapists are also under a professional, and I think moral, obligation to refer our patients to a colleague (or better, include a colleague in the ongoing relationship so as not to abandon our patients) when our affective reactions become an obstacle in the relationship. Note that affective responses often function as an obstacle in *any* relationship, let alone a therapeutic one. Furthermore, therapeutic models differ in approach to the other party in the relationship. Some see 'patients' (e.g., cognitive behavioral models) or just brains in need of chemical titration (psychopharmacological models), whereas others see 'persons,' (e.g., gestalt model). (I leave aside the manner in which Strawson specifically botches the psychotherapeutic model.) The point here is that taking up the objective view does

not straightforwardly dictate a 'non agency' *perception* or *stance* on the part of the therapist, as Strawson's use of the objective view stipulates our disposition. There is no uniform conception of 'agency' and 'self' across models of treating psychopathology. So, we should not take care that a 'she' not become an 'it' when diagnosed with a mental illness. When we depersonalize the mentally ill we may not only disrespect them in so doing, but worse, pave the way for further disrespect and even abuse. When we speak of type 2 type subjects in the Strawsonian language as objects to be 'controlled,' 'managed,' 'manipulated', we must not forget that mental illness, in addition to being a moral concept, is a concept that we put to use in our social and political institutions.<sup>49</sup> This is not always done with sufficient gravity. (In many respects, this language itself is morally abhorrent.) Suffice it to say that this discussion should serve as an antidote to our extolling the objective view as an appropriate response to type 2 agents, psychopaths included.

## 2.5 Conclusion

There is a third sense in which we might conceive detachment and this sense will open up new options that allow us to move beyond Strawson's stark distinction between the reactive attitudes and objective view. Here I only wish to introduce the idea and will return to further discuss its role in forming the response to psychopathic wrongdoing that I favor. Given that affect plays an inescapable role in the reactive attitudes and this underwrites some of the concern with having the

---

<sup>49</sup> See Ian Hacking (1999) for a discussion of mental illness as an interactive kind rather than as a natural kinds

reactive attitudes toward type 2 agents, then a key distinction may alleviate this concern. We can distinguish two types of *reactanda* (the things being reacted to) to which we direct our *reactans* (the reaction itself). The first *reactandum* is the *agent*. This is what most of the literature, including proponents of the common view, have focused on thus far in assessing the responsibility of psychopaths. Instead, we can, and should, direct our *reactans* (our affective reactive attitudes) to *behavior*. There are many gains to be made when we draw this distinction. By doing so we avoid a key concern raised by proponents of the common view (that we harm type 2 agents when we hold them responsible, which is only *really* done when the reactive attitudes are appropriately involved). Additionally, focusing on behavior as the appropriate *reactandum* allows us to maintain, or maybe even initiate, genuine interpersonal contact that begins to serve as an antidote to alienation, and maybe even to increasing the agency of ‘the other.’ Lastly, this distinction has the added benefit of being consistent with the way in which those with Antisocial Personality Disorder are *actually* diagnosed – and that is by their *behavior*.

Before closing this chapter I wish to note an important insight that emerges from the preceding discussion, which is that rather than trying to apply the standard responses to wrongdoing by ‘normal’ people and then apply that standard to inform our responses to abnormal people, like psychopaths, we should instead take a lesson from responses to abnormal wrongdoing and apply them more generally to all agents. We should take caution not to ever emphasize the reactive attitudes as having neither privileged insight into, nor an exclusive hold on moral responsibility.

## **Chapter 3: Levy's view**

### **3.1 Introduction: Levy's general view and its appeal**

Levy (2007a, b, 2010) provides the most current and well supported argument that we should not hold psychopaths responsible because they lack the capacity for moral responsibility and that blaming them harms them unjustifiably. His view represents an empirically updated version of both Watson's and Strawson's accounts (what we might call a 'folk psychological' account in which the moral idiot suffers from 'misfortune of formative circumstances') that includes cutting edge data from fMRI scans that allege to reveal neuro-anatomical abnormalities in the brains of psychopaths that explains their inability to be morally responsible agents.

After providing a brief account of Levy's view and its appeal, I engage four of his arguments supporting the conclusion that psychopaths are not morally responsible. First, Levy's prioritizing of the 'is responsible' question (as a rather stark disjunct: 'is the psychopath responsible or not?') to the 'hold responsible' (which, like Strawson is rather coarse-grained) is discussed as the backdrop to the empirical argument. Second, Levy's twofold value-free commitment, that both the concept of psychopathy is purely empirical and value-free and perhaps more provocatively that the question of the moral responsibility of agents is an essentially empirical matter, is challenged. Third, I criticize what I take to be an inversion of the responses to psychopathic wrongdoing called for by Levy, that we imprison psychopaths (perhaps indefinitely) so long as we refrain from blaming them. Last, I engage Levy's argument that psychopaths lack the capacity for moral knowledge

(another refinement of the ‘understanding’ condition advanced by Strawson and Watson).

The appeal of Levy’s view lies primarily in the idea that our moral responses should track the moral capacities of agents. His account is additionally valuable in supplying empirical research to provide insight into the capacity for moral responsibility. If a psychopathic agents lack the capacity for moral responsibility, and we can show this with empirical research, then the moral response of blaming psychopaths is both epistemically unjustified and perhaps, as Levy argues, is morally wrong too. The appeal of an evidence based approach to explaining the capacity for moral responsibility possessed by psychopaths lies, in part, as a welcome countermeasure to the tendency to censure and blame wrongdoers, perhaps especially the most inimical wrongdoers. Our understanding of moral agency, (and mental illness for that matter), is enhanced by evidence.

A brief analogy might help. On Levy’s view, to make moral demands of psychopaths is tantamount to demanding someone with a spinal cord injury to ‘run.’ It would be grotesquely ignorant of us to demand that quadriplegics will themselves to stand. Likewise, to make moral demands of psychopaths and then blame them when they fail to meet those demands, displays a similar ignorance capped off with a kind of self-indulgence that frequents blaming. And when we hear of the horrible childhood suffered by psychopaths like Robert Harris, there is something wrongful about adding to his misery by morally condemning him. Surely we should not add to the misery of psychopaths.

### 3.2 Causal responsibility and morally responsibility

The empirical evidence Levy provides is in support of a larger causal argument that psychopaths lack moral responsibility. Since psychopaths cannot cause their actions with specifically moral reasons, he argues (2008, 2010), they cannot be morally responsible agents. Hence, psychopaths lack the moral volition necessary for moral responsibility. Levy argues in favor the volitionist theory of responsibility in which to be responsible for an action, omission, etc., and agent must have sufficiently caused the action or omission by properly choosing to act. Per Levy's volitionist account, since psychopaths are not the proper causes of their selves, then they do not choose their reasons for acting in the way that grounds moral responsibility. Hence, on Levy's view, there appear to be only psychopathic *causes* of action (that emanate from the illness) in which psychopathic *reasons* for action are thereby pathological reasons.

(Interestingly, Levy appears to sweep up the attributionist account of responsibility – that an agent is responsible for a given act in so far as that act is attributable to *her* – in his volitionist account by maintaining that psychopaths do not have a proper self because it was not a self that was chosen by the psychopath. However, the 'choice of self' condition is not something that clearly separates psychopaths from the rest of us.<sup>50</sup>)

Given Levy's commitment to the view that psychopaths do not properly cause themselves it seems odd for him to admit elsewhere (Levy, 2007a) that

---

<sup>50</sup> Galen Strawson (1993) offers a compelling argument against the *causa sui* notion of self.

psychopaths are 'causally responsible for a large percentage of crimes,' (referencing Reznick 1997). Levy of course goes on to argue at length that psychopaths are not specifically *morally* responsible (2007a,b, 2008, 2010). Yet Levy, (2007b, p. 169) also remarks that he believes the question of the moral responsibility of agents can be largely answered 'without actually doing any moral philosophy.' If this is the case then we appear to be solely engaged in a question of whether psychopaths are the cause of their own actions. Since Levy admits that psychopaths are causally responsible for their actions – and this is supported by their engagement in instrumental aggression which requires calm, intentional, deliberate and planned actions – then the degree of agency and control psychopaths appear to possess over their actions makes it very difficult to see why they are not both causally responsible and morally responsible. The failure of psychopaths to intend morally good or worthy actions does not itself demonstrate that such failure is itself not blameworthy.

On Levy's view psychopaths are undeniably causally responsible for their criminal actions but are not specifically morally responsible for them. Levy also believes that the moral responsibility of a given agent is merely a matter of free agency, where moral theory need not be among our considerations. Hence, moral responsibility is about causing actions freely, and psychopaths are causally responsible for their actions in so much as they freely do them, but are not morally responsible for them. (The space between the causal freedom of action psychopaths possess and the moral unfreedom they likewise possess is, more so the size and significance of that space, is the primary domain of dispute over the responsibility of

psychopaths.) If moral responsibility is determined, in part, by causal freedom and psychopaths are causally responsible for their actions to a high degree (to which Levy consents), then they may also be, in *some* part, morally responsible for their actions. Causal knowledge appears to establish at least some sense of moral knowledge and moral responsibility.

Drawing from the cases, there is a sense in which both Harold and Lori are causally responsible for the damage that resulted from their actions. But neither Harold nor Lori intended to harm anyone so it seems incorrect to say that they are morally responsible for the harm that resulted. The gulf between the causal responsibility and the moral responsibility borne by Harold and Lori is *very* wide. Psychopaths on the other hand deliberately intend to cause harm to others with such a clear sense of causal control over their own actions that makes it very hard to see why they are not likewise morally responsible for the harm they cause. The gulf between the causal responsibility and the moral responsibility is much narrower because of the calm behavioral control exhibited in the signature behavior of instrumental aggression that psychopaths are known for possessing.

### **3.3 The epistemic condition for moral responsibility: understanding, reasons, and 'pathological' reasons**

The section below takes up the question of the role in which moral understanding informs the capacity for moral responsibility. Levy (2010) attributes psychopath's lack of moral understanding to a failure to *grasp* the force of moral demands. I'm not sure, however, that the way in which Levy's portrays the

psychopath's failure to grasp is accurate. I am equally unclear on what exactly Levy means by grasping a moral demand, in part because of the phenomenological nature of this condition. (Furthermore, why one must possess a high level of conceptual command, if this is what we mean by 'grasp', in order to be a proper target of the reactive attitudes is also unclear. Responses to wrongdoing can, and should, contain the attempt to promote an increased level of moral understanding by the wrongdoer.) I provide a sample of responses that reflect, or appear to reflect, various depths of understanding. Doing so affords the opportunity to examine the alleged differences between bad reasons, superficial reasons, and 'pathological' reasons. The latter among these, when held against the backdrop of a disease model of mental illness (and psychopathy), are viewed as causes more so than as reasons. So when we refer to 'pathological reasons' what it really meant is 'dysfunctional cause,' or 'neural failure,' or something similar. This is not always (and maybe only rarely) what has actually occurs as the madness isn't simply or easily revealed in the reasons given; it is also inferred by the listener, however correctly or incorrectly. Thus, as Jaspers forewarned, (and even Freud), that psychiatry contains an irreducibly hermeneutic element, particularly so as these crafts are actively practiced. This hermeneutic element follows us to the armchair as well depending on our portrayal of the psychopath.

Quite aside from assigning moral understanding as a condition for moral responsibility is the question of what moral understanding actually amounts to. A strict read is to require that one be morally autonomous in order to possess moral understanding. But (full) moral autonomy may be an overly strict standard for

moral responsibility. That is, one can be less than autonomous a still morally wrong someone.

Moral autonomy is not so much a condition for moral responsibility as it is an idealization of a robustly moral agent. But who among us are good *practicing* Kantians? Besides, autonomy may not be the moral ideal that we believe it to be. On my view, Levy's account of moral responsibility requires that agents possess a level of moral autonomy that is too high as well as a depth of moral understanding that is too deep, which leaves us with overly narrow the population of morally responsible agents. There are many examples in which the apparent failure to grasp a moral reason is *itself* grounds for blame.

Consider the following reason, 'I killed him because it was Tuesday.' On Levy's view, this reason is evidence of a psychological disease. (To 'clarify' what I mean by disease here, in part, is the causal role disease is assumed to play as a specifically potent reason for why we should think an agent is not responsible. This is not strictly true, however. Responsibility is not silenced by disease, including those mental illnesses that are argued to belong under the same conceptual umbrella as physical diseases. In fact, even the presence of paradigmatic physical disease – say, cancer – reveals the *increased* responsibility that a dying person has to her family.) Since the reason is deemed as incoherent, it is inferred to be a mad reason rather than a bad reason. To see this, let's examine the argument:

1. If an agent provides an odd/abnormal/incoherent reason for her action, then her reason arises from a psychological ailment. (Thus, it is deemed to be no reason at all.)
2. The agent provided an odd/abnormal/incoherent reason.
3. Hence, the agent is ill. (And is, this, not really an agent at all.)

Of course, the conclusion does not necessarily follow. Perhaps the agent was under orders to kill someone on a specific Tuesday, or maybe even one among several Tuesdays because the victim had a specific routine that afforded the assassin her opportunity. This counterfactual counterexample shows that this argument is invalid. So, the inference from bad reasons to mad reasons is something to we should actively guard against. Recall Levy committing this pathologizing error with his interpretation of welfare maximizing responses to the trolley problem, in which utilitarians and psychopaths appear to provide pathological moral judgments. Drawing the inference from 'bad reason' to 'mad reason' is weak in this case of 'Tuesday' being the reason cited for acting. Although 'it was Tuesday' is an odd, unclear and perhaps unreasonable reason, it is far more coherent than, say, glossogonal utterances as a response. A glossogonal 'reason' does make the inference to 'mad reason' far more plausible than the reason 'it was Tuesday.' Suffice it to say, the clash of intuitions about when a bad reason is better explained as a mad reason is central to the problem of enlisting mental illness as an intuitive guide to understanding the nature of moral responsibility. On my view, moral philosophers are far too liberal in such attributions and the burden should be much

higher (for both epistemic and moral reasons). It is easy to understand the appeal, however. Bad and really ignorant reasons psychopaths give for acting violently (e.g., 'she said I'm fat') *do* tempt us to explain away such lack of insight to madness.

Nonetheless, attributing madness to an individual (by virtue of the reasons they give for acting) carries moral consequences, (the insane are divested of civil liberties).

Let us examine a list of reasons for acting (violently) that provide insight into levels of competence. Whilst examining these different reasons we must keep in mind that what we are ultimately after is a standard of moral understanding (as it is held internally by an agent) that gives insight into the level of responsibility possessed by an agent – and not whether an agent 'is' or 'is not' responsible as if there is no middle ground. We must then hash out where this normative standard is to be drawn. On Levy's view one must *grasp* moral concepts to be morally responsible; and he picks his examples of reasoning to suit his conclusion. This may be too high of a standard, however (let alone how phenomenological thick this concept is). Perhaps the standard should be lower, requiring a more basic conversational competence short of a full cognitive command.

Examples of 'reasons' for acting (violently):

1. 'I didn't kill *him* but I did kill the dragon.'
2. 'I killed him because it was a Tuesday.'
3. 'I killed her because she said I was fat.'
4. 'I killed him because the coin landed on tails.'
5. 'I mugged her because she was easy prey.'

Each reason in the set reveals increased normative (and psychological) competence and yet all four fall short of an ideal appreciation for the force of moral norms. The first is a response we might hear from a person under the influence of hallucinogenics or in the throes of a severe psychotic episode. Yet there is *some* sense of moral accomplishment this person has for having killed a dragon, thus the moral competence is not absent or even the problem. The problem is that the psychotic killer thought that the person she killed was a dragon; hence her global disorientation to reality is the problem. The second reason was discussed above, and shows that we cannot simply infer a failure of moral competence with this response, not at least without more context of the situation. We need to be clear that this second reason is *not* indicative of psychopathic responses. The third response is an *actual* paradigmatic psychopathic response. It is a rebarbative reason no doubt, (more so when it is uttered gleefully). It also appears to be a superficial reason and certainly a reason that fails to grasp the force of moral norms.

Reason four is adapted from Cormac McCarthy's book *No Country for Old Men*, in which a psychopath 'decides' who he will kill or let live by the flip of a coin. Now, one might quickly point out that one cannot properly make a moral decision by flipping a coin. And even if one alleges to 'decide by the coin', then such an agent is not properly reasons responsive in the way that grounds responsibility. Neither of these of these objections are that convincing, however. It may well be that the psychopath (meta)decided that for his own amusement he would introduce an arbitrary system to 'decide' who to kill or not, and (unlike in the book) make his victims aware that their life hangs in the arbitrary balance for his morally insulated

amusement (since he can't control the outcome of the toss). Why the structure of this decision to harm others fails to qualify as a reason for acting that is subject to the moral appraisal of others is perplexing to me. It seems more like *psychopathic wholeheartedness*. The insertion of the coin by the psychopath is intentional so that he can *rationalize* (in the psychoanalytic sense of the term, not the epistemic sense) his evasion of responsibility. His ill will toward others does not simply evaporate into the decision-making structure than he deliberately instituted. It is manifest in his carrying out the verdict called for by the coin. This is one example of instrumental aggression (deliberate and sangfroid as opposed to reactive and impulsive aggression) that was identified in chapter 1.

Reason five, "I mugged her because she was easy prey," is perhaps the *most* paradigmatic psychopathic on the list of reasons above. It is a practically rational reason but a morally rebarbative one all the same. The fact that the mugging psychopath does not place any worth on his victim, and only on what he can possibly divest her of, does very little by way of mitigating the wrongness of his action.

Which on this list is a *mad* reason? I suspect that only the first on this list represents an obvious exemption from moral responsibility, (provided we are certain we are not being conned or duped by a con artist feigning psychosis – something a psychopath is clever enough to pull off, by the way). Many on this list fail to be anything remotely on the level of an agent fully appreciative of the force of moral demands, but it does not follow that 2-5 are not blameworthy reasons because they are mad reasons.

To summarize the picture before us: depending on which among these reasons we think are bad reasons and which are mad 'reasons', will determine our assessment of the moral responsibility of the agent giving these reasons. This in turn will instruct our normative responses to that agent, and whether we should blame her or detach from her, or impose various sanctions on her. The concern I raise here is that Levy's view requires us to lump far too many bad reasons in with the mad reasons. To support my criticism, Levy's view (actually, cf. 2010) commits us to including welfare maximizing reasons *as mad reasons*. I find this to be a dangerous move – and what we might call the 'pathologization of bad moral reasons.' The allegation of when a bad reason is better understood as a mad reason deserves much more careful treatment it ordinarily receives.

The nature of moral standards may also be a source of disagreement between me and Levy. Perhaps an analogy will help. If a logic student learns that affirming the consequent is a valid argument form, just because she fails to grasp her mistake and continues to routinely affirm the consequent, it does not render her arguments valid or otherwise free from scrutiny. Whatever conclusions she draws from her arguments are invalid regardless of her grasp. To reorient the rules of inference to the student's failed logic is to pervert the rules to suit an individual (or even for the entire group of routine consequent affirmers). If the student believes that affirming the consequent is a valid form because her logic teacher taught her this, then perhaps he is responsible for the error of her ways, but he can't be responsible for all her invalid beliefs. When the student leaves the classroom and enters the world, her invalid reasoning renders her prone to judgment, maybe even ridicule from

others and that is a responsibility that her logic teacher *can't* bear, unless she is present for her student's ridiculing, grasps the criticism and owns her role in the problem. At best, the student may be *ajustified* for her invalid reasoning. Her reasoning is bad, wrong and invalid but we may still attune the manner in which we set her straight. The specific response to the failed logic student is the normatively contested area of debate; the standard of validity is *not*. I suspect the very same thing is happening with the debate over psychopaths. Only, rather than a failed 'logic agent', psychopaths are failed moral agents.

But are the standards of morality sufficiently like the rules of logic for this analogy to be accurately drawn? As for a comparative metaphysics of the two, I am not qualified to comment. But metaphysical parity of logic and morality is not the point. The example above could just as well refer to those individuals believing the earth is flat. Here the failure of conceptual *grasp* does nothing to nullify the epistemic error. So, is the psychopath morally *ajustified* when he assaults and kills? No. The consequences of affirming the consequent are relatively low compared those of psychopathic killing. In both cases there has been a failure to meet the basic standards, but in only one of those results in victims. The fact that psychopathic actions produce terrible moral consequences, even though the psychopath places no moral weight – aside from his own amusement – on the outcome makes it very hard to see why we should exempt psychopaths from responses that contain moral sentiment.

The upshot of this section is that we cannot assume that a failure of action entails a failure of *grasp*, nor then can we rule out the possibility that psychopaths

*flout* moral norms. Because psychopaths *do not* appear to weigh moral values does not entail that they *cannot*. Nor do we need them to for that matter. What we need from the psychopath is for him to refrain from his violent and aggressive behavior and, in so much as this occurs, he may be morally improving. (Later in chapter 4, I entertain the objection that a ‘reduction of symptoms’ is not what we have in mind by moral improvement. I will argue that we can make sense of the idea.)

Furthermore, the over emphasis on providing adequate causal accounts of psychopathic actions distracts us from taking up the task of setting the norms for properly responding to psychopathic wrongdoing. Since the task of putting to rest the allegedly factual question whether or not the psychopath *is* responsible remains an unresolved debate, it makes some sense, if not at least to relieve ourselves from the strain of inquiry, to take up the question of how we should respond to psychopathic wrongdoing. In the following section I summarize and then criticize the response to psychopathic wrongdoing that is suggested by Levy, specifically that he fails to acknowledge the cost, perhaps even the harm that his response entails.

### **3.4.1 Levy’s inversion**

For as much attention to detail that Levy pays to the (allegedly) factual question of whether or not the psychopath *is* responsible, his account is not equally careful with respect to the question of how to respond to psychopathic wrongdoing. Levy does not attend closely to the question of how we should hold psychopaths responsible because he is focused on first answering what he takes to be the factual question of whether or not the psychopath ‘is’ responsible. On my view, the

response to psychopathic wrongdoing that does emerge from Levy (as an afterthought) is morally problematic. Thus, I focus in detail on Levy's suggested response to psychopathic wrongdoing below. I first provide some context with respect to three senses of 'holding responsible' to situate Levy's view and my criticism of it.

Fisher and Tognazzinni (2013) claim that for an agent to be morally responsible for an act or omission results in 'opening' them to the reactive attitudes of others, and perhaps as well, to 'suffering and sanctions.' To be 'open' in this way can have (at least) two senses. The first sense is that one is open to responses of others. The opening is provided to the potential blamers, regardless of whether they avail themselves of the opportunity. The second sense refers to the sense in which the offender is herself open, or receptive, to the reactive attitudes or suffering and sanctions when they are enforced. (Note: This second sense is, I suspect, what Watson means by the second 'face' of responsibility, 'answerability.' I am not entirely clear if the first sense of 'open' identified above is what Watson means by the first 'face' of responsibility, 'accountability.' I think it is close in meaning but needs added to it the account of responsibility known as 'attributionism', in which case a person is responsible only if we can attribute the act or omission to *her* – to her person, self, etc.) It is this second sense of opening, in which the offender is either, broadly speaking, conciliatory for the wrongs she is being called out for committing, or rather, she remains steadfast, defiant or defensive in owning up to the wrong she did. When all goes right, offenders apologize for the wrong they have done and the consequences other suffered as a result; here responsibility is 'taken'

or 'owned.' This does not occur with psychopaths, however, as this openness is not to be found. (Despite the evidence offered by Levy, we still do not know *why* psychopaths are this way. We only know *that* psychopaths are this way.) The nature of this absence of openness is what is so polarizing about psychopathic responsibility. To some this lack of openness represents a willful refusal to acknowledge responsibility for wrongdoing arising out of an entrenched intransigence. Yet to others, specifically Levy, this absence of openness shows that psychopaths lack moral sense, moral understanding, and likely the neural structure necessary for moral responsibility.

The focus of this section is primarily on the first sense of openness, in which case we must give an account, or justify, our responses to psychopathic wrongdoing. In chapter 3 we distinguished between *holding* the reactive attitudes and *expressing* those attitudes in response to psychopathic wrongdoing. We have not so far said much about a third type of response in the form of negative consequences, including punishment, detention, or more generally, to bring suffering and sanctions to the wrongdoing psychopath. As we shall see, on Levy's view, only the last on this list is a justified response to wrongdoing psychopaths. After summarizing Levy's view I offer some objections.

### 3.4.2 Levy's response to wrongdoing psychopaths

Consider the following claims made by Levy:

"In my view, an agent is morally responsible for an action just in case he or she deserves to be treated in certain ways in virtue of her having performed that action." (2007a, p.133)

"Holding psychopaths responsible imposes *moral* costs. It requires us to treat agents in ways they do not deserve."  
(2007b, p. 134)

"We can excuse the psychopath for her wrongdoing, without being committed to releasing her to offend again." (2007a, p. 136)

I respond to each individual claim and then criticize what I call the Levy inversion (although he is not alone in holding this view).<sup>51</sup>

With respect to the first claim, I simply want to note how this claim connects to another claim of Levy's that we can resolve the question of moral responsibility empirically and without reliance on moral philosophy.<sup>52</sup> This claim shows, however, how much Levy relies on the specifically moral notion of *desert* in forming his response to wrongdoing psychopaths. Since I will argue later (chapter 4.5) in this dissertation that the common view places too much emphasis on history (as a *reactandum* alleged to be of paramount consideration), and since the moral notion

---

<sup>51</sup> Strawson (1993), Watson (1993), Fisher and Ravizza (1998), Fisher and Tognazzini (2013), Benn (1999)

<sup>52</sup> Levy (2007b) fn. 5

of *desert* shares the same historical focus, it follows that my view likewise resists emphasizing an historic laden notion of desert in judging the moral responsibility of an agent. A critical analysis of desert would take us too far afield, but suffice it so say that the spirit of my objection is that future considerations should factor into our concept of desert as well. (Levy's use of *desert* is itself under-described.) It is not clear, however, that psychopaths deserve to be detained indefinitely nor that they deserve to be approached merely from the objective view. Perhaps psychopaths do deserve to be blamed but we are not yet in a position commit to this claim.

Claim two and three paired together is, in my view, rather astounding. To repeat Levy, "[h]olding psychopaths responsible imposes *moral* costs," (author's emphasis). And, "[w]e can excuse the psychopath for her wrongdoing, without being committed to releasing her to offend again." Why this pair of claims is astounding arises from Levy's other claims that psychopathy is both a developmental and terminal disease. On Levy's view, to make moral demands of psychopaths is wrong because moral responsibility is an "option that will always be closed off to her," (2007b, 137). Levy takes great pains to establish that psychopaths possess nothing in the way of moral capacity. I myself cannot reconcile the two competing notions.

Levy assumes that to hold persons responsible in general is to exact a moral cost. In fact, it is also assumed that we have only successfully held an individual responsible if that person accepts or takes responsibility for their wrongdoing. This imbues the practice of holding responsible as a kind of emotional commerce. There

are serious problems with such an approach. (See Nietzsche's *On the Genealogy of Morals* for attacking our moral practices because they center on resentment and the desire to make others *pay*.) Furthermore, it assumes something that is rather contrary to Levy's insistence that psychopaths thoroughly lack moral knowledge and moral sense, and that is if we are imposing a moral cost on the psychopath then he must be *paying* emotionally, or in some other sense. For if the psychopath's moral account is bankrupt, as the common view insists, then there is no moral currency with which to pay. Hence, no moral cost can be exacted from an agent whose agency is already devoid of morality. This worry cannot, on pain of consistency, be a problem for Levy's view. But it can represent a problem for my view, however.

Since I believe that there remains something within the psychology of the psychopath that *can* impose a moral cost, then my view is immediately faced with justifying whatever moral cost is actually imposed on the psychopath. I think we can. Given his relative insensitivity, we are less in a position to morally harm the psychopath than we are to morally harm a non-psychopath when we affectively express blame toward them (and not their behavior or their reasons). In fact, don't we desire to impose a moral cost on the wrongdoing psychopath? His paying up represents an ideal result.

On my view, Levy must forfeit one of the following two claims. Either he must acknowledge that the psychopath is not as morally blind as he portrays (and he must admit, on pain of evidence, that psychopathy is not terminal; most simply 'get better' around age forty). Or he must acknowledge that blaming psychopaths

harms them by exacting a moral cost. For if there is no cost to exact then there is no moral currency to be found in the economy of the psychopath's psychology.

It may be that we can morally harm non-moral agents without them being aware of the harm that is done to them. We morally harm animals when we factory farm them, but to say that they are aware that the harm being done to them is a specifically moral harm seems odd. Perhaps all harms done in the slaughterhouse are moral harms. To the point, if as Levy suggests, psychopaths are non-moral agents whom we can nonetheless morally harm, can we morally harm psychopaths by indefinitely detaining them? Or, by confining them to solitary housing units even within prison? Mentally ill inmates (particularly the violently ill) occupy more than their fair share of solitary units. The evidence of the psychological harm of solitary confinement speaks to the moral problem of detention, which does not appear to be among Levy's consideration. I cannot make sense of the view that imprisoning or otherwise detaining someone – even for a finite period, let alone *indefinitely* – does not exact a moral cost. In fact, a prison sentence often carries with it the intent, by *raison d'état*, to exact a moral cost. And not only does imprisonment exact a moral cost from individual offenders ('the guilty must pay!'), it exacts *all sorts* of costs (financial, psychological, emotional and moral) from the family members of offenders as well as society at large.

Perhaps we justify detaining the psychopath by drawing a quarantine analogy. The quarantine analogy is problematic because it commits us to the prevention of psychopathic wrongdoing rather than as a response to it. Can we justify preventive detention of psychopaths? Consider the neurologist researcher

who discovered that images of his own brain revealed that he too is a psychopath – only one whom has never behaved psychopathically. If we were serious about the quarantine analogy we would be obligated to round up all individuals whose fMRI fit the profile and detain them, only this time our refraining from blame *is* appropriate. Detaining psychopaths who have done no wrong is not morally justifiable, however. (Besides, the evidence that ‘non offending’ psychopaths exist appears to present serious counter-evidence to the common view’s account of the psychopath – as an individual *determined* to harm others. I think this suggests that, despite all the evidence from fMRI, the continued inclusion of psychopathy among the list of mental illnesses should be seriously reconsidered.)

Levy’s inverted approach sees blaming the psychopath as a more severe moral offense than indefinite detention. But we typically we think that affective expressions of blame interfere less with a wrongdoer’s autonomy than physical incarceration. It is one thing to affectively blame someone for her wrongdoing, (whether that takes the form of censure or a harsher condemnation, for instance), but it is quite another to restrict a person’s liberty of movement by isolating them in prison. And since psychopaths are so interpersonally (and morally) calloused, it is very difficult to see how a verbal expression of blame can do *any* harm to a psychopath, let alone a *moral* harm. Even if it turns out that blaming psychopath does indeed morally harm them, and since we want to harm non-psychopaths when we blame them, then psychopaths and non-psychopaths appear have more in common than the common view acknowledges.

I want to pause momentarily in order to briefly query what is 'in' a blame that is so worrisome to some philosophers. Clearly Levy places considerable weight on the act and attitude of blaming. Though he says little about the actual content of blame, we might infer that it is infused with the paradigmatic attitude of resentment and it is the resentful attitude of the offended party that is alleged to morally harm the offending party. Nietzsche warns us that resentment is most harmful to the resenter. (Perhaps this is why we appear so ready to give it away.) Consider Watson's (1996) preferred response to psychopaths: 'we can fight the psychopath, and call him a 'son of a bitch' among other things, but we can't reason with him in the way we can morally responsible agents.' I suppose Levy would not include fisticuffs and swearing among the list of morally harmful responses. Quite aside from the common view's limited range of moral harms, this view begs the question that a person is reasonable in advance of the demands of morality. Racists and misogynists are unreasonable too and we see them as paradigmatically failed moral agents whom are nonetheless blameworthy for the state of their moral character as well as for individual actions. Once we remove medical content from the label 'psychopath' that we began to apply to the persistently aggressive wrongdoer (Cleckley, 1955) then we have before us even *more* of a failed agent than a racist, though psychopaths do come in racist and misogynist varieties. In fact, the very first case of a psychopath documented by Cleckley is of a misogynistic domestic abuser that he wrongly labels a 'psychopath' (because he felt the field was in need of this clinical title).

### **3.5 Conclusion**

In this chapter I have criticized Levy's thesis that psychopaths are not in fact responsible and accordingly, should not be held responsible. I began (in chapter 1) by epistemically disputing Levy's confidence in the empirical evidence, and furthermore, revealed how and where Levy's analysis mistakes, perhaps substitutes, factual evidence for normative evaluation (specifically in his assessment of the trolley problem) and how a controversial reason is thereby reframed as a pathological reason. I then turned to the question of how we should respond to wrongdoing psychopaths and criticized Levy's response as unduly inverted.

## Chapter 4: Beyond the common view

### 4.1 Introduction

I argued in the previous chapter that, with respect to psychopaths and other type 2 agents, the common view recognizes only one way of going wrong when we (mistakenly) attempt to hold these agents responsible and that is *that we* inappropriately hold the reactive attitudes toward them. I also argued that this results in proponents of the common view overlooking the moral downside of retreating to the default objective view. Furthermore, the resources that the objective view provide is assessed from the common view exclusively in terms of the offended party. It likely does not seem like such a great resource to those being managed, manipulated and controlled, however. Proponents of the common view assume that the value of the objective view as a resource specifically for the offending party comes by way of providing a kind of protection from blame and resentment. (This is Strawson's position by default; Levy argues specifically for this claim, however.)<sup>53</sup> The value of the objective view is not analyzed independently or in much detail, and its value is assessed *only* in how it is *not* like the reactive attitudes. I also argued that the reactive attitudes and objective view have more in common than Strawson recognizes and then raised some concerns with the objective view, including the tendency to perpetuate, and even reinforce, the alienation of type 2 agents. The starkness of Strawson's distinction, between the reactive attitudes and the objective view, specifically with respect to their

---

<sup>53</sup> Levy 2007a, 2007 b

application to the 'moral idiot,' requires us to seek new options for responding to psychopathic wrongdoing that, ideally, make the best of what both the reactive attitudes and objective view have to offer.

Accordingly, this chapter is an attempt to conceptually map out some new options for responding to psychopathic wrongdoing. This task is not merely to seek out new options but also to explicate the rationale for a better response to psychopathic wrongdoing than offered by the common view. Furthermore, this chapter attempts to provide something missing from the common view and that is to articulate our responses in more detailed and operationalized form, for moral language is often bewitching. For example, consider the phrase, 'we should not hold the psychopath responsible.' This is far too nebulous to provide much at all in the way of informing our responses, neither internally (intrapsychologically) nor externally (in terms of our behavior). To be fair to philosophers arguing to exempt psychopaths from responsibility, the detailed accounts of what psychopaths are alleged to lack in their moral psychology are discussed in great and insightful detail. Not so, however, when it comes to details regarding our responses to psychopathic wrongdoing. What philosophers do provide, often as afterthoughts, are two general tactics that we first, refrain from blaming psychopaths, and second, that we lock them away indefinitely. However, silence and isolation are two morally debatable policies. Perhaps there are morally superior options to holding psychopaths responsible.

To begin, I return to a distinction drawn at the close of chapter 2 between *reactans* (our reactions) and *reactandum* (what our reactions map onto). The

preceding discussion of Strawson afforded an initial but not complete survey of *reactans*, with respect to the broad distinction between the reactive attitudes and the objective view. To fill in this discussion with more detailed options, two broad candidates of *reactans* are offered below: the *holding* of attitudes as intrapsychological states that are distinct from the *expressing* of attitudes as interpersonal exchange. (Discussion of a third response to psychopathic wrongdoing, the enactment of consequences, including punishment, is postponed in order to focus more closely on distinguishing holding and expressing the reactive attitudes.) After discussing two candidates of *reactans*, I turn to mark a distinction of *reactanda* that is overlooked by the common view. This will highlight both the shortfall of the common view while simultaneously revealing an option for holding psychopaths responsible in a way that addresses a legitimate concern that arises from the common view. If philosophers want to do right by the psychopath, as common viewers plainly claim to wish, we need an option that calls for more than the common view conclusion of detachment and isolation.

With a more detailed structure of both *reactans* and *reactanda* in place we can see both where and how the common view is limited. (Strawson and Levy use the notions of holding and expressing the reactive attitudes interchangeably, making it unclear, for instance, how our resentment should, if at all, show itself to the offending psychopath. Watson, on the other hand, provides interesting detail about how our attitudes might unfold as we learn about Richard Harris' deeds and then his personal history. Albeit improved detail from Watson, I will argue that our reactive attitudes can be more nimble than even he acknowledges.)

## 4.2 Two types of *reactans*: holding and expressing the reactive attitudes

When organizing types of *reactans* the most general distinction, following Strawson's schema, is drawn between the reactive attitudes and the objective view. (I have already argued that this distinction is too starkly applied to psychopaths. Also, I raised concerns about using psychopaths as a tool designed to illustrate this distinction.) In this section, however, I am more interested in further distinctions of reactive attitudes that can be drawn on a middle level, above that of distinguishing particular, or token, attitudes such as resentment or gratitude.

We can distinguish *holding* the reactive attitudes from *expressing* them in our practices of holding others responsible. (We will ignore the third category of consequences for the time being.) When we speak of *holding* a reactive attitude we are describing an internal (intrapsychological) state that is some kind of fluid combination of doxastic attitude coupled with an emotional attitude. Our belief that we have been wronged corresponds with some emotional attitude, and vice versa. (Unfortunately, resentment is the paradigmatic attitude frequently cited in the literature, but as noted, may be a poor standard to guide us in our theorizing about the nature of responsibility). Though often a bare reflection of the internal attitude we hold, the *expression* of the attitudes *can* be markedly different than the *holding* of an attitude. Perhaps our expressions *should* be different than our holdings. (I will first explain the ways these can differ before defending the claim that they should differ, specifically with respect to psychopaths.)

When drawing this distinction, in both theory and practice, we must maintain vigilant awareness of the gap that occurs between the reactive attitudes we hold internally and how they are revised when expressed. We must balance the practical gains we seek in our expressions with remaining genuine to the attitudes underlying them. (Perhaps the genuineness to which I refer maintains what Strawson calls the 'vital thing' that is contained in our reactive attitudes that points us to moral responsibility, and that which he takes to be unaccounted for when taking up the objective view and considering the social utility of the attitudes.) We need not be overly concerned, however, that the wider the gap between holding and expressing the reactive attitudes necessarily entails that we are moving out of the domain of moral responsibility. Accordingly, the distinction between holding and expressing the reactive attitudes is specifically important for informing our responses to psychopathic wrongdoing.

A further distinction with perhaps even more importance is one that is drawn by mental health professionals between affective and non-affective reactive attitudes.<sup>54</sup> My use of 'affective' here follows the clinical meaning of the term as describing an individual's emotional state and the relative passion or depth of that state as it is embodied. For example, 'the anhedonic exhibit a flat affect.' More to the point, we have a tendency to engage in affective blame, and this comes in relatively charged degrees. I want to pause at this point to note that on my understanding of the common view, it is this particular subtype of *reactans* that

---

<sup>54</sup> This distinction is common in clinician use. See Hannah Pickard's (2008) use of the terms 'affective and non-affective blame' in her approach to working with individuals diagnosed with personality disorders.

common viewers take to be the exclusive domain of the language of moral responsibility. So, when common view instructs that we ‘should not hold psychopaths responsible,’ what is meant is that we ‘should not blame them.’ On my understanding it is this specific domain of *expressing affective reactive attitudes* that we are warned to avoid. I return to this point shortly to crystallize my objection to the common view that this is an overly narrow approach to holding responsible. We might issue instead, a *non-affective expression* of a given attitude. To imagine this, we must first purge ourselves of the paradigmatic ‘go to’ attitudes of blame and resentment, otherwise the incoherent idea of non-emotional resentment prevents us from considering alternatives. Rather, issuing moral demands can be done absent the heightened level of affective content that Levy and others assume *must* accompany all form of ‘holding responsible.’ Our concern here is not so much the underlying reactive attitude that is actually held as it is with the manner that that attitude is expressed *qua* a moral demand.

Summarizing these distinctions then, we have the general distinction between the reactive attitude and the objective view; under the reactive attitudes we have the internal holding of the attitudes that are distinct from external expression of the attitudes; and under each of these we have affective and non-affective subtypes. This yields the following possible *reactans*: internally held affective reactive attitudes, internally held non-affective reactive attitudes, externally expressed affective reactive attitudes, and externally expressed non-affective reactive attitudes. (To be thorough we would have corresponding divisions under the objective view: internally held affective objective attitudes,

internally held non-affective objective attitudes, externally expressed affective objective attitudes, and externally expressed non-affective attitudes. Common view philosophers will hasten to point out two items on this list that are, in their view, incoherent: the internally held non-affective reactive attitudes and the internally held affective objective attitudes. But all this charge reveals is the narrow importance common viewers attach to the role of emotion in how we justify our holding others responsible. Thus, the role of emotion in our practices of holding other responsible is a significant source of dispute between my view and the common view.)

Before leaving this section, I want to examine a viable candidate from the list above that provides a sensible and morally acceptable approach to holding psychopaths responsible, the non-affective expression of the reactive attitudes. This candidate *reactans* will shortly be coupled with the candidate *reactanda* (behavior) that will emerge at the conclusion of the following section.

Given that psychopaths are so difficult to identify and distinguish from other thoroughly bad agents, when we are assaulted, and thus offended by a psychopath, we are likely to form an internal affective reactive attitude that is unfiltered in our expression of that same internally held affective reactive attitude. To be clear, it is the affective component, in both the holding and expression of the attitudes, with which common viewers appear to be most concerned. At best, the more levelheaded among us might be able to uncouple an internally held affective attitude enough to express that attitude non-affectively. To be more specific on this point, if we can internally detach from the intensity of the affect involved in the attitude

*enough* to express a however slightly filtered version of that attitude, then we may have a response option to psychopaths that common viewers will find attractive. But this internal affective detaching is, no doubt, quite difficult to achieve, especially when we are reacting to the aggressive behavior that defines the psychopath. It may be, however, that in the very attempt to non-affectively express the affective attitudes we hold, we will in turn aid in the internal detaching, or reducing the affective of, the internal state we hold. We might think of this attempt as a measuring of release that is less intense than naked catharsis.

So, the candidate *reactans* we have before us is the non-affective *expression* of the reactive attitudes. Now, what of the targets of these reactions, the *reactanda*?

### 4.3 Two types of *reactanda*: agents and behavior

When we form a reactive attitude, (momentarily ignoring the distinctions just discussed above), to what is it exactly that we are reacting? For brevity, let us refer to the ‘thing to which we are reacting’ as the *reactandum*. For understandable reasons, the common view assumes the agent is the obvious and proper *reactandum*. This is also generally true of the wider literature on moral responsibility.<sup>55</sup> However, there are additional *reactanda* that can be accounted for,

---

<sup>55</sup> Consider George Sher’s (2008) claim, “If anything about the concept (responsibility) is clear, it is that what any given agent is responsible for is a function only of what *he himself* has done or failed to do,” [emphasis added], p. 99. Levy (2010) notes two distinct theories of moral responsibility are generally recognized in the wider literature: attributionism and volitionism. The former focuses centrally on the freedom of the agent and whether or not an act is directly attributable to her, whereas the latter focuses on the freedom of an individual’s

and though we might identify a lengthy list of these (reasons, choices, personal history, etc.), I wish to focus on behavior as the *reactandum* that is particularly relevant to the debate whether or not we should hold psychopaths responsible. The general case for shifting our reactive attitudes to the behavior of the psychopath rather than the psychopathic agent himself is first briefly described. This is followed by some observations of the *reactanda* as accounted for in the common view, and noting why these are alleged to attract our reactive attitudes and their suspension. I then defend the behavioral *reactandum* approach against some objections.

In order to see the value of the shift from steering our reactive attitudes away from an agent and, in place, toward her behavior, we need to recall a core concern held from the common view that we should not hold psychopaths responsible (specifically, that we should refrain from blaming them) because doing so both harms them and places an unreasonable demand on them. The common view objects to blaming psychopaths because they are alleged to be agents incapable of understanding even basic moral obligations (Strawson, Levy) and because blaming them would harm psychopaths in a way they do not deserve (Levy). These objections raised by the common view arise from a prioritizing of the *diseased agent*

---

choice in so acting. With respect to volitionism, we might say that the relevant *reactandum* is the freedom status of a choice a given agent makes, but the assessment of the freedom of choice will result in *reactans* that are directed at the agent. Levy (2007, 2010), though a professed defender of volitionism, is particularly prone to collapse volitionism into attributionism when discussing psychopaths. For instance, he claims that psychopaths are not responsible because, owing to their moral disease, they do not choose to be the person they are, thus, they cannot intend to act morally, and thus, are not responsible for the things they do.

and her personal history as the *reactanda*. But an agent's mental illness and her history are not the only relevant *reactanda* that can, and do, inform our reactions. We do, and should, include behavior and an agent's future to the list of *reactanda*. Perhaps behavior should be the *reactanda* that warrants prioritizing. Let's look at the case for doing so.

Suppose for a moment that the common view is correct and that when we express reactive attitudes to psychopaths we harm them. Suppose also that we are wary of alienating psychopathic agents by taking up a purely objective view toward them (and furthermore, suppose that we are wary of the conclusion that we simply lock psychopaths away indefinitely). Here is where the option of prioritizing behavior as the more apt *reactandum* gains its appeal. For, if our reactive attitudes are centered on behavior then we avoid a direct attack on, or criticism of, the character or agent herself. It is one thing to condemn a person and quite another to condemn behavior. Although this may be true generally, it seems to apply even more so in cases of emerging agents, including psychopaths. Character, including self-perception of our own character, is much more difficult to change or modify than behavior and behavioral patterns, especially when we metaphysically overcommit to the idea that behavioral change can only follow character change. For, when we comment on character (as revealed in the reactive attitudes) we are alleging to comment on something constitutive of an agent, something that is more deeply rooted than behavior, more so patterned behavior, which is alleged to reveal one's character. If, however, we diminish the commitment to direct our reactive attitudes at an agent's character and prioritize behavior instead, then we have more

reason for optimism that moral improvement can occur by way of behavioral modification. Rather than seeing character, or to be precise - diseased character - as the source of behavior, perhaps we need to acknowledge that character and behavior actually work in a kind of gravitational tandem. This approach also helps avoid the conclusion that psychopaths suffer from an incorrigible disease of moral character, for which there is no treatment.

When we prioritize behavior as the more apt *reactandum*, then the distinction between *holding* and *expressing* the reactive attitudes previously discussed gains salience. It just may be a brute fact about us that when we hold a reactive attitude internally that we do so with the *reactandum* of the offending agent firmly in mind. It does not follow from this, however, that when we *express* the held attitude that the expression cannot be redirected from the agent to direct reference to the offending behavior rather than the agent herself, her character.

It may be that we need to help psychopaths believe that they do not have permanently damaged moral characters and that wrongful behaviors need not, as a matter of Pavlovian reflex, beget a lifetime of wrongful behavior. Recall Robert Harris' confession in his final hours that when he killed the two boys that he knew then he had made his choice to go to hell. Unlike Watson's experience of reactive attitudes toward Harris, it is this confession that is the first, and perhaps the only, time in reading his story that I felt compassion for him. (Whether it arose from a religious commitment, or a moral metaphysical commitment, or an underlying belief that his formative years were so bad that he was psychologically ruined, or what have you, what is saddest about Harris' story is his idea that he was eternally

condemned for his behavior. I think this suggests that there was guilt somewhere deep within Harris' moral psychology.) If we take a closer look at Watson's reactive attitudes to Harris, we can see a key source of the debate over whether or not we should hold psychopaths responsible.

As Watson describes the flow of reactive attitudes we feel when we read about Harris' case, the antipathy we feel *toward Harris* when we hear about his execution of the two boys and his ensuing fantasy of impersonating a policeman in order to tell the parents of the young boys about their killing so that he may be entertained and amused by their grief turns to sympathy when we read about the horrible abuse Harris suffered in his childhood. Now, I must admit, my reactive attitudes do not make this shift from antipathy to sympathy. To paraphrase Wittgenstein regarding the gestalt duck-rabbit illusion, the duck does not turn into a rabbit for me, even when hearing about Harris' history. Perhaps I am overly censorious. The account provided by the common view here that would urge me to shift my *seeing as* victim instead of victimizer goes as follows (and we will note the *reactanda* that the common view emphasizes).

For example, the shift from an attitude of antipathy to one of sympathy as describe by Watson, occurs with a shift of *reactandum* from the murdering adult Harris to the child victim Harris. Phenomenologically, for Watson, the antipathy *felt* is altogether eclipsed by sympathy. The problem, as I see this response, is that it assumes much like the duck-rabbit example that we can *only* take up one attitude to the expense of other competing attitudes. Not everyone, however, is so skilled at making the duck-rabbit switch, nor making the switch many times and nimbly so. I

think this shifting precisely what is called for in how we should approach holding psychopaths, like Harris, responsible. So, rather than allowing our reactive attitudes to categorically shift in *reactandum* from Harris the bad adult to Harris the unfortunate child, (thus, the shift in *reactans* from antipathy to sympathy), we must balance these two considered *reactanda*, and perhaps even balance these two with a third consideration, the future Harris.

Taking stock then, yes, when we learn of Harris' terrible childhood we feel sympathy for him given the abuse he suffered, and when we learn of the horrific murders he commits, we feel moral disgust toward him. But none of these observations strictly entail that the reactive attitudes we *express*, despite the complexities of our mixed attitudes, should not make specific reference to his behavior. Nor must we go about (attempting to) hold psychopaths like Harris responsible by only considering how our reactive attitudes are informed by historical concerns.

To further sum up the common view account of the things to which we react, the assumption is that the reactive attitudes must always be directed at the psychopathic agent and the historical and psychological considerations that (somehow) define his lack of moral agency. Accordingly, the common view sees the psychopath as an agent whose moral development was arrested in his youth and can never be restored in any capacity. From this it follows, according to the common view, our reactive attitudes do not apply to him. However, as I have argued in this section, the expression of our reactive attitudes can, and should, target behavior rather than agents. And when we express reactive attitudes in response to

wrongdoing psychopaths, those expressions need to comment directly on the (in this case, immoral) behavior so as to avoid giving psychopaths the impression that their moral character is so rotten that they too believe they should be isolated and locked away indefinitely.

#### **4.4.1 The common view's narrow *reactans-reactandum* pairing**

Now we are in a position to better see the narrow sense of both the *reactans* and *reactandum* involved in 'holding responsible' that the common view has in mind. When the common view instructs us that we should not blame the psychopath, what is meant is that, in terms of our *reactans*, we should not express affective reactive attitudes, and in terms of *reactandum*, that our reactans are assumed to be specifically to be reserved for agents. And when take a closer look at the common view's preferred *reactandum* in the case of the psychopath, we see a diseased agent (maybe not even an agent at all) and the agent's past history. And, given this narrow *reactandum*, the common view argues, it is wrong for us to respond with the narrow *reactans* of affective blame.

Of course, we need not commit to such a narrow sense of either the *reactanda* or the *reactans*. Or, if you like, we need not commit the common view's narrowed sense, but rather commit to another narrow paired sense of *reactanda* and *reactans*. My suggestion then, is that we re-conceive what it means to hold a psychopath responsible. We do this first, by acknowledging the substantial objections to the very idea of a psychological disorder *qua* moral disease because

the standards for the disease are normative and not medical. Second, we take up an alternative *reactans-reactanda* pairing, specifically, non-affective reactive attitudes, the expression of which specifically target behavior rather than the agent's character.

What is so troubling about psychopaths is that, given the commitment to the narrow sense of *reactans-reactanda* pairing, assumed by the common view, our typical practices work with most people. To be more specific, our customary practice of merging our affective holding of attitudes with an affective expression of those same attitudes (and with matching magnitude) targeted at agents that is so effective in our uncontroversially normal lives, has zero effect on psychopaths. And since our traditional modus operandi of holding responsible utterly fails with psychopaths, the common view infers from this that we have on our hands a thoroughly morally diseased quasi-agent. (Even more oddly, we infer something like Levy, that psychopaths are physiologically lacking a conscience. This, of course, is a rather egregious category mistake.) But why should we think that our traditional practices of holding responsible, (in the narrow *reactans-reactandum* combination just discussed), is a standard that, when *it* fails, reveals that the problem *must be* with the non-responding agent, and never with the standard practice itself? We should not. Perhaps the case of psychopaths forces us to acknowledge that the commitment to (justified) affective blame is a poor standard for holding responsible.

It is also fair to say that the history of philosophy gives us many different accounts of how we should go about responding to the moral wrongdoing of others.

Often when we seek to address a moral mistake made by others the range of interventions available are wide. Kant, for instance, might nudge the conscience of the wrongdoer whilst respecting her autonomy and refraining from being a moral busybody.<sup>56</sup> No philosopher, however, was busier than the gadfly Socrates. Unafraid to 'bite into the flesh' and harm the stupidity of the wrongdoer, Socrates would put his interlocutor to shame if necessary, and where shame fails, eschatological appeals are thrown in for extreme measure.<sup>57</sup> (Socrates is notably unable to put Calicles to shame in *Gorgias*. In fact, this very strategy of putting to shame for the goal of moral improvement is what – I suspect – Nietzsche takes to be the core error of all post-Socratic philosophy, hence his own 'campaign against resentment.')<sup>58</sup> Plato is even more heavy-handed, particularly with the incorrigible temple robbers for whom execution awaits for those who morally fail even once more.<sup>59</sup> (The Greeks took the concept of moral pollution seriously indeed.) But the steadier hand of Aristotle guides us seek out a response to offense that displays our own virtue as well as promoting the development of virtue in the character of the wrongdoer.

---

<sup>56</sup> Thanks to Adam Cureton for pointing this out.

<sup>57</sup> The exchange between Socrates and Calicles, the third interlocutor in *Gorgias*, is particularly noteworthy to this discussion. Calicles is the character that most closely resembles a highly intelligent, instrumentally aggressive, psychopath that we find in Plato (aside from the incorrigibles discussed in *Laws*) because Socrates' old successful trick of putting his interlocutor to shame utterly fails with Calicles.

<sup>58</sup> While all of Nietzsche's writings might be described as a campaign against *ressentiment*, he first acknowledges this motivation within his revaluation of values in the book, *On the Genealogy of Morals*.

<sup>59</sup> Plato's *Dialogues*, *Laws IX*

The range of available interventions appears to shrink when we are forced to address not just a single moral mistake, but patterns of wrongdoing. Habitual wrongdoing, not unlike major single offenses, elicits in us a tendency to narrow our psychological responses to deeply held affective states, thus blinding us to alternative options for responding. (It is of course easy to criticize this delimiting tendency from the detached comfort of the armchair. Such is the enterprise of the 'cool contemporary style' of philosophy noted by Strawson.) Thus, the ensuing frustration we have with habitual wrongdoing arises because our customary response, which is to say our reactive attitudes, appear to have no affect. From this, we then infer that habitual wrongdoing is really the more thoroughgoing incorrigibility, again, *because* our reactive attitudes do not result in changes to behavior or attitude we expect. (Incorrigibility is then further explained as being caused by a disease.) But this could be wrong and it may be that our limited ranges of responding are just as much to account for the lack of affect in responding to psychopathic wrongdoing. Nor is it the case that when our reactive attitudes fail to have the desired affect on the wrongdoer that we can infer that the wrongdoer is not a responsible agent. This is to confuse moral responsibility with moral responsiveness and to confuse the problem of shamelessness as residing solely in the wrongdoer.

So, when we ask the psychopath what led to the killing of his wife (instead of asking, 'why did *you* kill your wife?'), and his reply is that 'she said I was fat,' we do not ask rhetorically and with affective charge, 'what is the matter with *you*!?' Nor do

we exclaim, 'You should be ashamed!' This is what I suspect the common view has in mind by 'holding responsible.' Nor do we bite our tongue, disengage and detain him indefinitely. Rather, we might respond by saying something like, 'since killing is typically considered morally wrong, and very strong moral reasons are required to justify killing, help me understand why your wife's saying that you are fat is a strong moral reason that justifies kill her.' This response carries no affective charge, and neither does it attempt to heap shame on the psychopath himself. What it does is reinforce a moral boundary and suggest that a particular act of his appears to run afoul of moral standards, and invites him to give an account of his reason for acting, (however appalling those are likely to be). For now, this response marks an improvement from the common view. It is however, not complete, as it is a rather abstract and hyper rational style response. But the absence of affect is part of the cost of responding to the psychopath whilst maintaining some hope for promoting his moral responsibility. Beyond this response, however, there is also a case to be made that even affective attitudes have a place in our exchange with the psychopath. For if we want to promote his moral responsibility, and we remain committed to the role of affect in our moral relationships, then the psychopath will need some assistance in placing more (or *any*) weight on the affective states of others.

#### **4.4.2 Is gratitude prohibited too?**

The common view is silent on the appropriate role, if any, of the positive reactive attitudes (or more, the *affective* expression of positive reactive attitudes).

Should we refrain from affective expressions of the positive reactive attitudes, like gratitude (or good will in the Aristotelian sense?), toward psychopaths? If we take a strong common view stance, then we should refrain from holding and expressing all reactive attitudes to psychopaths, both negative and positive alike. Strawson's view steadfastly commits us to a purely objective view; hence, even the positive reactive attitudes are inappropriate to hold toward psychopaths. Levy's view might not be as strong, however, and allow for expressions of gratitude (when called for, because, presumably, psychopaths are capable of token acts where gratitude is called for by the beneficiaries of the act).

Beginning from Levy's view that we should not blame psychopaths because doing so harms him, then how we are to extend this 'principle' to the positive attitudes is not clear. We should also note here that Levy's principle too appears to hinge on the outcomes of our response to psychopaths; we should not blame him *because* it harms him. (I am not the only party committed to weighing the consequences of our various responses to psychopaths.) Suppose then, by extension, that our expressing gratitude to a good-doing psychopath *benefits* him in some way. If so, and if we are committed to producing certain effects in the psychology of psychopaths, then it appears that we should express gratitude to them. If we do justify expressing gratitude because of the effect it has on psychopaths, then a key premise in Levy's argument must be false, namely, the claim that psychopaths have no moral sense. For if we acknowledge that gratitude psychologically (maybe morally) benefits the psychopath in some way, then there must be some, however minute, moral sense on which gratitude gets its purchase.

There is something then within the psychology of psychopaths (within their agency) that we have to work with. Furthermore, it does not matter if he lacks 'full moral sense', he possesses enough for us to express the full range of attitudes to him, however nuanced in delivery these are to be expressed.

#### 4.5 Objections

I consider two objections to close this chapter. The first is that, per the common view, I am wrong about the source of agency alienation in the case of the psychopath. The second is that my own emphasis on behavior as the more appropriate *reactandum* commits my view to another version of agency alienation that I aim to avoid.

The common view proponent will no doubt object that the source of alienation of type 2 agents that I bemoan is not to be found in the reactive attitudes of the offended party, and that, rather, the source of alienation is the mental illness! And this mental illness is what *causes* the psychopath to be alienated from his own agency, which then results in alienation from others. In short, the mental illness *causes* all of the intrapersonal and interpersonal alienation. But, this is a poor argument for several reasons.

First, I have already argued extensively against the causal disease model of psychopathy in chapter 1, so I will not repeat that argument here. I wish only to point out that the disease model of psychopathy is wrong for two reasons: there is insufficient empirical evidence that psychopathy lies in anatomy, and, that the

criteria by which one is diagnosed with psychopathy (Antisocial Personality Disorder) are *entirely* behavioral and normative. These two objections to the disease model support my argument that the question 'is' the psychopath responsible or not is to be answered either in the affirmative, or that it remain an open question. I then suggested that to better inform how we should respond to psychopathic wrongdoing we should begin with the assumption that the question of psychopathic responsibility (inasmuch as it can ever be established as fact) remains open. But this seems odd, for how are we to begin to form a policy that is entirely neutral with respect to whether or not we are dealing with a responsible agent? Perhaps we really are unable to begin our deliberations without making some assumptions; that seems to be a plausible observation. All I am suggesting here is that we exercise some suspension of prior belief, even if this falls short of what D.T. Suzuki calls the 'beginner's mind,' where there are more, not fewer possibilities.

Second, even if the causal story that is assumed in the common view was in fact true, (that the disease of psychopathy is the causal fountainhead of all ensuing alienation), it does not follow that we, as the responding offended party, should perpetuate the alienation. The defense that the disease alienation precedes the alienation compounded by the reaction is no defense at all. As reactors, we are in some sense *responsible* for reacting in such a way that does not perpetuate alienation, and maybe we should even respond in ways that begin to reduce the alienation. One way to do this is to maintain the inherently human reactive attitudes so that the offending party, even when these are psychopaths, experience genuine human exchange. (Of course, this assumes that we define our humanity as

genuine expressed in the reactive attitudes. I am not sure, however, that inasmuch as resentment *feels* genuine to us, that it is an attitude reflecting what is most genuine about humanity.)

Regarding the second objection that my view is inconsistent by de-prioritizing agency as a *reactandum*. This is an important objection, but one that has a response. Recall that I criticized Strawson for two ways in which he suspends agency with respect to the psychopath. The first is the idea that psychopaths intuitively lack agency, and second is the idea that, owing to his illness, he lacks agency such that we should rather see him as something merely to be controlled or manipulated. My own call to steer our reactive attitudes toward behavior rather than the agent is a suspension based on a practical strategy that aims to prevent agency from being completely written out of the picture, with respect to the psychopath. This practical strategy simply subordinates, but does not extirpate, metaphysical concerns regarding the agency of the psychopath. In fact, when we direct our reactive attitudes toward (psychopathic) behavior rather than (psychopathic) agents, we keep do so with the agent firmly, yet tacitly in mind. After all, we want to respond in the most respectful way possible and we want her to modify her behavior, and for the better morally. By emphasizing behavior, and behavioral modification, we at least, however slightly, keep open the possibility of 'making' psychopaths more morally responsible.

## 4.6 Conclusion

The fact that psychopaths appear to be impervious to blame shows only that our customary habits of holding responsible by way of blaming do not work and that, rather than giving up altogether, we should revise our practices. Blaming psychopaths is impractical, but this does not show – nor entail – that psychopaths are thus not morally responsible for the wrongs that they do. Psychopathic imperviousness suggests that our old narrowly conceived practices of holding each other responsible solely conceived as blaming need to be re-evaluated. It also shows that blame no longer corners the conceptual market of moral responsibility. And rather than entirely writing psychopaths off as non-agents to be isolated and detained indefinitely, we should seek out new methods of demanding basic moral expectations.

## Conclusion

In philosophy there are rarely, if ever, conclusions. The discussion remains ongoing. But I hope in this dissertation to have thrown some doubt on a prevailing paradigm in several interconnected ways.

Psychopathy is a value-laden concept such that its inclusion among the value-free medical diseases is doubtful. So far, empirical research gathered on the brains of psychopaths has not yet established the anatomical origin of their wrongdoing or of the disorder. As yet, there is no corresponding lesion of the brain that provides a causal explanation of psychopathic behavior. Aside from commenting on the state of the empirical evidence and the level of epistemological credence we lend to that evidence, I also hope to have shown with a detailed a specific example of how the insistent search for an empirical cause of psychopathy can lead to carelessness with respect to the presence of values that are *in* the concept of psychopathy. I hope to have raised awareness to the smuggling of facts into an inherently value-laden concept of psychopathy. This is not to be taken as a commitment to a Luddite conception of moral responsibility. Levy is to be commended for bringing empirical facts and a scientific worldview to the difficult question of the moral responsibility of psychopaths. My objections to Levy, I hope, serve as a warning and reminder that even when we take up a scientific view, (including what Strawson calls the 'objective view'), that moral and social values remain present in both the concept of psychopathy as well as sorting out how we should respond to his wrongdoing. For if we desire a morally sensible response to psychopathic wrongdoing, then good

evidence must be matched with good values. In one sense, my dissertation can be viewed as an effort to bring values back up to pace with the focus on burgeoning empirical data gathered on psychopath's brains. The debate over the responsibility of psychopaths is incomplete without both facts and values involved in our considerations.

In addition to showing the ways in which values remain central to the concept of psychopathy I also attended closely to how moral philosophers suggest we respond to wrongdoing psychopaths. I argued that we need to seek out better responses than simply locking psychopaths away indefinitely. I also offered alternative ways in which we might hold psychopaths responsible by demanding basic moral expectations of them and that one way we can do so is by way of the non-affective expression of the reactive attitudes.

This brings me to some of the ideas to which I can now turn my attention. These ideas include both forward thinking about the more general problem of mental illness and moral responsibility that will be better served with a good history, or genealogy, of the emerging tension between disease and resentment.

In chapter 4, I argued in favor of directing our reactive attitudes (preferably expressed non-affectively) at wrongful behavior rather than at an agent's character. In as much as such an approach is designed to promote the moral responsibility of psychopaths, however small the behavioral gains, it is on my view a far more sensible response to psychopathic wrongdoing than the common view approach to detach, detain and isolate. My approach remains committed to expressing moral demands from psychopaths without maintaining a commitment to mere censure

and judgment typically associated with blaming. The next and perhaps most challenging task is to confront psychopath's reasons – or, more specifically and ideally, to help psychopath's confront *their own* reasons.<sup>60</sup> Since reasons-giving and reasons-receptivity is central to our understanding of what it means to be morally responsible agents, (and confronting reasons is central to the practice of holding others responsible), it follows that we want to hold psychopaths responsible in the reasons domain. This is a very difficult conceptual task.<sup>61</sup> In this dissertation I was only able to object to the tendency of philosophers (especially Levy) to insist that psychopaths can only produce mad reasons and never bad reasons. I revealed how Levy (2010) committed the pathologizing, bad-to-mad, move in his treatment of the trolley problem (in which controversial moral judgments are deemed pathological moral judgments). What Levy and others of a similar mind owe us is an argument demonstrating why psychopaths can only have mad reasons rather than bad

---

<sup>60</sup> My own work with psychopaths (primarily with the nascent version, Conduct Disorder) leaves me with the suspicion that psychopath's possess rather exquisitely robust ego defensiveness (in the language of the psychoanalytic tradition). The defiant refusal to admit to wrongdoing displayed by psychopaths seems to me to stem more from a need to psychologically self-protect than from a completed ignorance of the rules of both society and morality. It may be that the diminished violence that comes with aging psychopaths reveals a loosening of defensiveness as much as it shows an emerging respect for others – or evidence of subsiding symptoms of a moral disease.

<sup>61</sup> It might even be dangerous to gather better empirical data. I suspect that the empirical evidence that we have gathered on psychopaths is deeply flawed – specifically methodologically. All of the empirical studies have been conducted in prisons with guards serving as proctors. Since we know that psychopaths want to please authority figures (once detained), then we know psychopath's responses to moral questions are designed to please the authoritative questioner. Someone psychopaths know to have no leverage over their prison sentence must conduct these studies so that we can separate genuine responses to moral thought experiments from ingratiating responses.

reasons. A simple reference to mental illness as the 'cause' is thus far insufficient. The first chapter of this dissertation can be viewed as an antidote to pathologizing the human condition of psychopathy by denying the tendency to identify a mad reason as one that originates from madness. Such circularity is not convincing. Furthermore, the social and moral norms that define the behavioral criteria for psychopathy make it very difficult to accept the view that we are dealing with a bona fide medical disease with this disorder. If moral madness is a disease it is so only by analogy. I find this analogy strained to the breaking point such that we need to entirely reconceive the condition and concept of psychopathy. And we need not do so simply so that we can indulge ourselves in cathartic affective expressions of blame at his character. Because we want to morally respect psychopaths we need to seek out better responses to psychopathic wrongdoing. We should do so because it is the right thing to do.

Throughout this dissertation, I've found myself wavering between saying that issues related to 'holding responsible' should be prior to issues concerning 'is responsible' and saying only that both warrant equal attention. I realize that I have not demonstrated the priority of 'holding responsible' to be true. I would like to explore a whether or not a solid argument can be offered for the priority of holding responsible. I am also highly suspicious that whatever tacit assumptions we have in mind about how holding other responsible is supposed to be understood inform, consciously or otherwise, our understanding of the 'is responsible' issue. And, as a corollary to my claim that both facts and values must be involved in our understanding of mental illness and moral responsibility, so too must we include

both the 'hold responsible' and 'is responsible' questions in considering the broader *gestalt* of moral responsibility.

The facts and values approach that is particularly pertinent to psychiatric ethics is, I think, pertinent to the emerging field of neuroethics. The work of Bill Fulford (1995 and more) shows us that the principles-based account of bioethics (also known as the 'Georgetown mantra') in which facts are first settled before ethical matters are taken up does not translate to good psychiatric ethics because the concepts of psychiatry are for more value-laden (and 'non-descriptive') than in the case of paradigmatic physical illness. In this respect, Fulford's work can serve as advance warning to any future neuroethics involving psychiatric disorder to not repeat the Georgetown mistake. A key concern I have with Levy's approach to the psychopath and neuroethics in general is this very mistake. Good psychiatric ethics requires close consideration of both facts and values.

This dissertation is focused narrowly on the specific mental illness of psychopathy. I am interested to further investigate how my arguments here (specifically the argument for redrawing the reactans-reactandum pairing) have to tell us about moral responses to people with other forms of mental illness and even to normal adult agents. Many normal people possess various cognitive and affective defects that fall into a subclinical range in a way that falls significantly short of the, I think oft idealized, 'responsible agent' of contemporary moral theory.

Complex issues in philosophy of science lie at the core of philosophy of psychiatry. Some of the issues discussed in this dissertation may serve as a contribution to the ongoing methodological debate between the natural science and

the human sciences. This has direct bearing on the ways in which the practice and treatment of mental health should operate. The tendency for psychiatry to view itself and present itself as a natural science runs a very dangerous risk of neglecting the values and ethics of a practice that remains inherently human and irreducibly normative.

With respect to my suggestion that we rethink reasoning from the *normal* holding responsible paradigm (in which we justify blame and resentment) to the *abnormal* (which leads us to the conclusion that the mentally ill are not morally responsible) – my worry is that blame and resentment may themselves represent a pathological response and, as such are not desirable qualities on which to build theories of moral responsibility.

In this dissertation I was unable to address ethical problems relating to social and legal policy, as psychopaths are involved. To predict ethical problems on the horizon of empirical research on psychopaths, there is a lurking challenge relating to so-called neuro-prediction (of aggressive behavior) and how this evidence can and should be considered in the legal domain. Since we already view psychopaths as quasi-human (analogies to wild animals and viruses – to be quarantined – are commonplace in discussions about psychopaths) then there is little impetus to protect the civil liberties of such individuals. And given the severity of problems that psychopaths present, it may be well positioned conceptually to inform our policies more generally toward mentally ill offenders. Likewise, there are very difficult ethical problems relating to scarce resource allocation. For, if psychopaths are merely rabid humans then we ought not devote any scarce resources to them,

either in the way of treatment or for detention. Perhaps, however, we can make better use of existing resources and likewise, arrest the tendency to categorize such persons among the beasts. Furthermore, we must recognize our moral failing (without shaming ourselves) as a society that produces psychopaths. One way to address this problem is to invest more resources to developing treatment approaches that target early intervention. An indication of success might be measured longitudinally and in a generation we will be (if all goes well) discussing how the natural reduction of psychopathic symptoms reduced from age forty to age thirty. But we must first make an auxiliary moral commitment to take up such a task. We must decide if psychopaths are worth the effort.

As for the need for the tension that we feel between the disease concepts of mental illness and moral responsibility, I think we first need a history of the emergence of this tension. I think a good place to begin is with Nietzsche's engagement with Socrates, in which Nietzsche urges us to see Socrates (and all 'moral wrongdoing') as an expression of sickness. Crucially, Nietzsche advocates this approach as part of his own campaign against resentment – more specifically against morality that is based on resentment – and sees the roots of this view as emanating from the psychology of Socrates. This 'story' of offsetting blame with sickness carries though, I think, into the methodological debate between the natural sciences and the social sciences in which we (still) see the fields of psychology and psychiatry attempting to straddle. The point of painting the picture of our intellectual history relating to madness and moral responsibility is to seek out a notion of responsibility that divorces itself from association with blame without

defaulting to the disease model of mental illness either. The disease model represents an overcorrection to the error of a blame-based morality, and rather ironically shares with blame-based morality, a failure to promote responsibility. This failure that is shared by the disease model of mental illness and a blame-based morality of resentment, does not, on my view, represent our best understanding of responsibility. A better understanding of the nature of responsibility requires that we move beyond disease and resentment.

I wish to close this dissertation by sharing an insight from my time working clinically with young men diagnosed with Conduct Disorder and believed to be on the fast track to psychopathy. I am also in debt to Levy for reminding me of this with his suggestion that we can no longer understand the psychopath, or moral responsibility for that matter, solely from the armchair and that we must also learn from the laboratory. I think we should take Levy seriously and even extend the reach beyond the laboratory and even outside the walls of the mental health clinic. Some insights are shy and must be taken by surprise. As if by surprise and arising from necessity of circumstance from a defeated clinician, I found empathy in those young men - more importantly they found empathy in themselves - in a place as just as least expected as it is obvious: hanging from a mountainside. Here we saints and scoundrels find common feeling.

## **Bibliography**

Adshed, G. (1999) "Psychopaths and other-regarding beliefs." *Philosophy, Psychiatry and Psychology*, volume 6, number 1, pp. 41-44, Johns Hopkins Press

Benn, Peirs (1999) "Freedom, Resentment and the Psychopath" *Philosophy, Psychiatry and Psychology* volume 1, number 6. Johns Hopkins Press

Blair, J., D. Mitchell, and K. Blair (2005). *The psychopath: emotion and the brain*. Malden, MA: Blackwell.

Blair, J.D. (2007) "What emotional responding is to blame it might not be to responsibility" *Philosophy, Psychiatry and Psychology*, volume 14, number 2, p. 149-151 Johns Hopkins Press

Brunswick, Egon. (1938/1952) "The Conceptual Framework of Psychology" in Neurath, O., Carnap, R., Morris, M. (eds.) *Foundations in the Unity of Science* Volume 1, p.659-760. Chicago: University of Chicago Press.

Boorse, C. (1997). "A rebuttal on health." In: *What is disease?* Ed. J. Hunter, and R. Almeder, 1-134. Totowa, NJ: Humana Press.

Cartwright, Cat, Fleck, Uebel (2008) *Otto Neurath: Philosophy between Science and Politics* Cambridge: Cambridge University Press

Charland, L.C. (2004) "Moral Treatment and the Personality Disorders" in *The Philosophy of Psychiatry: A Companion* Ed. by Jennifer Radden. Oxford University Press.

Cleckley, H.C. (1955) *The Mask of Sanity*, St. Louis: The C.V. Mosby Company

Cushman, F. and Young, L. (2009) "The psychology of dilemmas and the philosophy of morality." *Ethical Theory Moral Practice* 12:9-24. Springer.

Deigh, J. (1996) *The Sources of Moral Agency: Essays in Moral Psychology and Freudian Theory*. Cambridge University Press.

Doidge, N. (2007) *The Brain That Changes Itself* New York: Penguin

Edwards, C. 2009. Ethical decisions in the classification of mental conditions as mental illness. *Philosophy, Psychiatry and Psychology* 16, 1:73-90.

Elliot, C. (1996) *The Rules of Insanity: moral responsibility and the mentally ill offender*. State University of New York Press: Albany.

-----(2008) "Mental Illness and its Limits" in *Philosophy of Psychiatry: A Companion*, ed. Jennifer Radden. Oxford University Press: Oxford

Elliot, Carl and Harold, James (1999) "Travelers, Mercenaries and Psychopaths" *Philosophy, Psychiatry and Psychology* volume 1, number 6, pp. 45-48

Fischer, J.M. and Ravizza, M. (1993) (Eds.) *Perspectives on Moral Responsibility*, Cornell University Press: Ithaca

-----(1998) *Responsibility and Control*, Cambridge University Press: New York

Fischer, J.M. and Tognazzini, N. (2011) "The physiognomy of responsibility" in *Philosophy and Phenomenological Research*, V:LXXXII, No. 2, March.

Fulford, KWM (1989) *Moral Theory and Medical Practice*, Cambridge University Press

Gillett, Grant (1999) "Benn-Ding the Rules of Resentment" *Philosophy, Psychiatry and Psychology*, n.6 v.1, 49-51

---- (2010) "Intentional action, moral responsibility, and psychopaths" in *Responsibility and Psychopathy: interfacing law, psychiatry and philosophy*. Eds. Malatesti, L. and McMillan, J., Oxford University Press: Oxford

Haji, I. (2010) "The inauthentic evaluative schemes of psychopaths and culpability," in *Responsibility and Psychopathy: interfacing law, psychiatry and philosophy*. Eds. Malatesti, L. and McMillan, J., Oxford University Press: Oxford

Harre, R. (2004) "Setting Benchmarks for Psychiatric Concepts" in *The Philosophy of Psychiatry: A Companion* Ed. by Jennifer Radden. Oxford University Press

Hacking, I. (1999) *The Social Construction of What?* Harvard University Press

Holland, N. (1998) *The Madwoman's Reason: The concept of the Appropriate in Ethical Thought* Pennsylvania State University Press: University Park

Kendall, R.E. (1975) "The concept of disease and its implications for psychiatry." *British Journal of Psychiatry*, 127:305-315

Kennett, J. (2010) "Reasons, emotion and moral judgment" in *Responsibility and Psychopathy: interfacing law, psychiatry and philosophy*. Eds. Malatesti, L. and McMillan, J., Oxford University Press: Oxford

Levy, Neil (2005) "The Good, The Bad and the Blameworthy" *Journal of Ethics Social Philosophy*, volume 1, number 2, p. 1-16

----- (2007.a) "The responsibility of psychopaths revisited," *Philosophy, Psychiatry and Psychology*, volume 14, number 2, pp. 129-138. Johns Hopkins University Press

----- (2007.b) "Norms, Conventions, and Psychopaths", *Philosophy, Psychiatry and Psychology*, volume 14, number 2, pp. 163-170. Johns Hopkins University Press

----- (2007.c) *Neuroethics*, Cambridge University Press: Cambridge

----- (2010) "Psychopathy, responsibility and the moral/conventional distinction" in *Responsibility and Psychopathy: interfacing law, psychiatry and philosophy*. Eds. Malatesti, L. and McMillan, J., Oxford University Press: Oxford

Matravers, M. (2007) "Holding Psychopaths Responsible," *Philosophy, Psychiatry and Psychology*, volume 14, number 2, p. 139-142. Johns Hopkins University Press

----- (2010) "Policies, law and psychopathy: a critical stance from political philosophy." In *Responsibility and Psychopathy: interfacing law, psychiatry and philosophy*. Eds. Malatesti, L. and McMillan, J., Oxford University Press: Oxford

Morris, H. (1968) "Persons and Punishment," *Monist* 52: 475-501.

Nelkin, D. (2013) "Desert, Fairness and Resentment" *Philosophical Explorations* (Special Issue: Desert)

----- (unpublished) "Psychopaths, Incurable Racists, and the Faces of Responsibility" Talk delivered at the University of Tennessee, April 12, 2013.

Nichols, S., Vargas, M. (2007) "How to be fair to psychopaths", *Philosophy, Psychiatry and Psychology*, volume 14, number 2, pp. 153-155. Johns Hopkins University Press

Noguera, Anita R., (2000) "Mad, sad or bad. Moral Luck and Michael Stone." *Nursing Philosophy*, volume 1, p.158-168

Nordenfelt, L. (1987) *On the Nature of Health: an action-theoretical account of health*, Reidel Publishing: Dordrecht

Plato *Dialogues* Ed. John Cooper. Hackett Publishing

Pereboom, D. (2001) *Living without Free Will* Cambridge University Press: Cambridge

Pickard, H. (2011) "Responsibility without blame: empathy and the effective treatment of personality disorder" *Philosophy, Psychiatry and Psychology*, volume 18, number 3, pp. 209-223. Johns Hopkins University Press

Rafter, N.H. (1997) *Creating Born Criminals* University of Illinois Press: Urbana

- Rezneck, (1997), *Evil or Ill? Justifying the insanity defense*. London: Routledge
- Robinson, D. (1996) *Wild Beasts and Idle Humors: The Insanity Defense from Antiquity to the Present*. Harvard University Press.
- Scanlon, T.M. (1998) *What we owe to each other* Harvard University Press: Cambridge
- Sher, G. (2009) *Who Knew? Responsibility without awareness* Oxford University Press: Oxford
- Sinnott-Armstrong, Walter  
 ----- (2008) Oxford Lecture Leverhulme Series, no. 4. Audio recording:  
<http://media.philosophy.ox.ac.uk/uehiro/sinnottarmstrongleverhulme.mp3>  
 ----- (2010) et. al. "Neuroprediction, Violence and the Law: Setting the Stage,  
*Neuroethics* Springer Press  
 ----- (2011) "Personality Disorders and Responsibility" *Philosophy, Psychiatry & Psychology* volume 18, number 3, p. 245-248
- Smith, A. (2005) "Responsibility for Attitudes: Activity and Passivity in Mental Life" *Ethics* 115:236-271
- Strawson, P.F. (1962) "Freedom and Resentment", *Proceedings of the British Academy*, reprinted in (1993) *Perspectives on Moral Responsibility* Ed. John Martin Fischer and Mark Ravizza Cornell University Press: Ithaca
- Szasz, T. (1974). *The myth of mental illness*, 2<sup>nd</sup> ed. New York: Harper and Row Publishers.
- (1998). Commentary on "Aristotle's function argument and the concept of mental illness." *Philosophy, Psychiatry and Psychology*, 5, no. 3:203-207.
- Talbert, M. (2008) "Blame and responsiveness to moral reasons: are psychopaths blameworthy" in *Pacific Philosophical Quarterly*, 89, 516-535.  
 ----- (2011) "Unwitting Behavior and Responsibility" in *Journal of Moral Philosophy*, 8, 139-152.
- Thornton, T. (2007) *Essential philosophy of psychiatry* Oxford: Oxford University Press
- Wakefield, J.C. (1992). The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist* 47, 3:373-388

----- (2009). Mental disorders and moral responsibility: disorders of personhood as harmful dysfunctions. *Philosophy, Psychiatry & Psychology* 16, 1:91-99.

Watson, Gary (1993) "Responsibility and the limits of evil: variations on a Strawsonian theme" in *Perspectives on Moral Responsibility* Ed. John Martin Fischer and Mark Ravizza, Cornell University Press: Ithaca

----- (1996) "Two Faces of Responsibility" *Philosophical Topics*, volume 24, number 2

Williams, B. (1985) "How Free Does the Will Need to be?" *The Lindley Lecture*, University of Cambridge

Wolf, S. (1987) "Sanity and the Metaphysics of Responsibility" in Shoeman, F. (Ed.) *Responsibility, Character and the Emotions* Cambridge University Press: Cambridge

----- (1993) "The Deep Self View" in Fischer, J.M. and Ravizza, M. (Eds.) *Perspectives on Moral Responsibility*, Cornell University Press: Ithaca

## Vita

Matthew Ruble was born in Boone, North Carolina to the parents of Mary and Raymond Ruble. He is the younger brother of Melissa and uncle of Zachary, Madison, Hunter and Megan. He is the stepson of Kay Ruble and stepbrother to Macey, Maxwell and Molly. He attended Blowing Rock Elementary and continued to Watauga High School in Boone, North Carolina. Following graduation, he attended Appalachian State University as a scholar-athlete, playing on the varsity soccer team for four years and studying Philosophy, Religion and Psychology. He earned a Bachelor of Arts degree in Philosophy and Religion from Appalachian State University in May 1994. He went on to earn a Master of Arts degree in Human Development and Psychological Counseling from Appalachian State University in May 1997. After working as a mental health professional in an outpatient community mental health center he returned to graduate school to study the intersection of philosophy, psychiatry and ethics in the United Kingdom. He earned a Master of Arts degree in the Philosophy and Ethics of Mental Health at Warwick University, U.K. in September 2001. His Master's thesis is entitled: "Positive Psychology: A 'science' of the good life?" He returned to Appalachian State University where he worked with programs funded by the U.S. Department of Education designed to prepare first-generation college students for success in post secondary education. During this time he also taught philosophy as an adjunct instructor at Appalachian State University. In 2009 he accepted a graduate teaching assistantship at the University of Tennessee's Department of Philosophy allowing

him to pursue a Doctorate in Philosophy specializing in Ethics and Applied Ethics (especially the Philosophy of Psychiatry). During the summer of 2014 he was awarded the Prados Summer Dissertation Fellowship in Ethics allowing him to work full time on this dissertation.