

# Moral Responsibility & Mental Illness: A Call for Nuance

Matt King & Joshua May<sup>1</sup>

Forthcoming in *Neuroethics*

**Abstract:** Does having a mental disorder, in general, affect whether someone is morally responsible for an action? Many people seem to think so, holding that mental disorders nearly always mitigate responsibility. Against this Naïve view, we argue for a Nuanced account. The problem is not just that different theories of responsibility yield different verdicts about particular cases. Even when all reasonable theories agree about what's relevant to responsibility, the ways mental illness can affect behavior are so varied that a more nuanced approach is needed.

**Word count:** 7,897 (excluding abstract)

**Keywords:** free will, accountability, blame, excuse, psychopathology, mental disorders

## 1. Introduction

In the summer of 2001, in a small town outside of Houston, Texas, Andrea Yates drowned each of her five young children in a bathtub, one by one. Yates's psychiatrist had recently taken her off of Haloperidol, an anti-psychotic medication. In previous years, she had attempted to commit suicide and was treated for major depressive disorder. During her trial, Yates pleaded not guilty by reason of insanity, and the jury ultimately agreed. Her lawyer proclaimed the verdict a "watershed event in the treatment of mental illness," presumably because it promoted the idea that having a mental disorder can compromise one's free will and thus reduce one's culpability, even for terrible acts (Newman 2006).

Some vehemently resist such conclusions, however. Just over ten years later in Texas, Eddie Ray Routh was convicted of killing two men at a shooting range, one of whom was celebrated sniper Chris Kyle. A former marine, Routh had been diagnosed with post-traumatic stress disorder (PTSD) and schizophrenia. His counsel sought the insanity defense, but failed to convince the jury that Routh did not know his actions were wrong. The district attorney, Alan Nash, won the jury over, stating, "I am tired of the proposition that if you have a mental illness, you can't be held responsible for what you do" (Dart 2015).

---

<sup>1</sup> The authors contributed equally to this paper. The order of surnames is presented alphabetically.

When and how does mental “illness” or psychopathology sufficiently undermine one’s moral responsibility? This question figures heavily in legal discussions regarding both criminal liability to punishment and civil authority for private law decisions. The issue is also relevant to designing public health policy and to our ordinary practices of assigning praise and blame. Indeed, while philosophers have traditionally focused so intently on determinism as a threat to free will and moral responsibility, some have turned their attention to psychopathologies. Walter Glannon, for example, identifies “brain dysfunction” as the “real threat to free will” (2011, p. 69).

Our guiding question is how mental disorders affect responsibility for action. Our focus is only on *moral* responsibility, even though it’s intimately connected with other forms of responsibility, and even though we will sometimes draw on cases from the law. Discussions of this issue, particularly in the public sphere but also in academia, often assume that the relevant question is whether having a mental disorder generally mitigates moral responsibility. The usual approach is to consider people with some category of psychopathology—e.g. schizophrenia, autism, or obsessive-compulsive disorder—and determine whether the individuals are (or could be) responsible agents.

Against this approach, we argue that there is no generally supported inference from an individual’s having a mental disorder to any claims about that person’s responsibility. Indeed, we think the focus of inquiry should not be on possession of disorders but on the operation of symptoms in specific circumstances. Different disorders operate quite differently, and even within a given disorder, its symptoms don’t always have a singular effect on capacities relevant to responsibility. Importantly, while “symptom” is often used to refer to only the *undesirable* effects of an underlying disorder, another sense of the term includes any phenomena that are characteristic, indicative, or symptomatic of a condition, which includes unproblematic or even desirable effects (more on this later). After all, underlying dysfunction can remain while undesirable symptoms of a disorder are not manifest. In Alzheimer’s, for example, neurodegeneration typically begins long before the patient notices any symptoms, such as forgetfulness or confusion (Palop et al. 2006). As our discussion will make clear, it is the broader sense of “symptom” that is relevant to assessing whether an underlying disorder affects whether praise or blame is warranted in a given case.

Another major difference between our discussion and most extant accounts is that ours considers psychopathology generally, and so does not single out any particular disorder or set of disorders. Indeed, because the philosophical literature already centers on psychopathy and addiction (see e.g. Maibom 2008; Levy 2013), we deliberately focus on other cases that have received less attention.

Throughout the discussion we rely on the latest *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) as a guide to the identification of psychiatric disorders and their symptoms. Some researchers caution that the DSM is unlikely to group patients together in the most informative way for discovering facts about mental illness (e.g. Tabb 2015). A further complication is that many disorders have

high comorbidity with others, such that in any particular case there may be multiple disorders at work.

In response to such worries, the National Institute of Mental Health now employs a set of criteria for assessing research projects that departs from the constructs of the DSM. The new Research Domain Criteria require, for example, that researchers frame their projects as addressing certain psychological phenomena (e.g. anxiety, attention, loss) rather than specific disorders. A study of anxiety, for instance, may include patients exhibiting rather different kinds of disorders identified in the DSM (e.g. obsessive-compulsive disorder, generalized anxiety disorder, PTSD).

If this is right, however, it suggests that the clinical community is already moving toward something more like our account, which denies there are informative relationships to be drawn between mental disorders and responsibility. We counsel evaluation of responsibility on a case-by-case basis, and the relevant phenomena to focus on are symptoms (e.g. impulsivity, delusions, hallucinations, anxiety, psychological incoherence, melancholy, diminished motivation, memory loss, reasoning deficits, blunted affect, mania, difficulty focusing, loss of consciousness, powerful urges, and so on). We thus appreciate the movement to focus on symptoms in psychiatric research (although of course the DSM categories may remain useful in clinical settings). However, employing the DSM constructs would seemingly put standard treatments of the topic in a more promising position, suggesting an ontology to the disorders that might support the relevant generalized inference. Thus, while we do not commit ourselves to the DSM accurately carving nature at its joints, it is both useful for framing the discussion to come and charitable to our opponents. Ultimately, our argument doesn't turn on any dispute about categorization and diagnostic criteria.

Likewise, we take no stand on how to best understand mental illness or what properly belongs to it. We use the terms "mental illness," "mental disorder," and "psychopathology" interchangeably, whatever differences they might suggest. Given that we rely on the heterogeneity of mental disorders generally, it should come as no surprise that we are skeptical there is much to be gained in greater terminological precision.

We begin by characterizing the standard approach to mental illness and moral responsibility and contrasting it with our preferred account (Section 2). We then examine how one's theory of responsibility will affect whether or not the symptoms of a mental disorder excuse (Section 3). Next we discuss how, on any theory, symptoms can affect moral responsibility by bypassing or diminishing one's agency. However, we show that psychopathology affects agency in such a variety of ways that we cannot draw an inference about one's moral responsibility from the fact that one has a mental disorder (Section 4). Finally, we consider special cases in which symptoms might actually enhance one's responsibility, which raises a final complication for the standard approach (Section 5). We conclude with a more nuanced account of the purported relationship between psychopathology and moral responsibility, and we briefly discuss some implications of this approach (Section 6).

## 2. The Naïve View

Merely labeling a mental condition a “disorder” strikes many people as at least typically mitigating some forms of responsibility (Edwards 2009). Psychiatrists will often describe mental disorders to their patients in the same ways as non-mental illnesses, such as diabetes, precisely because it suggests the disorder is out of one’s control (Arpaly 2005).

Philosophers also commonly treat the possession of a mental disorder as an excusing condition. Galen Strawson, for example, identifies as paradigm constraints on one’s freedom the manifestation of mental disorders, including “kleptomaniac impulses, obsessional neuroses, desires that are experienced as alien, post-hypnotic commands, threats, instances of force majeure, and so on” (1994, 222; quoted in Meynen 2010). Similarly, when discussing cases in which individuals are simply exempt from being held accountable for their actions, R. J. Wallace includes as “accepted exemptions” cases of “insanity or mental illness” (1994: 165), along with childhood, addiction, psychopathy, posthypnotic suggestion, and the like. Daniel Levy (2003) even names more specific disorders—e.g. PTSD, Tourette’s, schizophrenia, alien hand, bipolar disorder, and obsessive-compulsive disorder—and considers them all “maladies” of free will (cf. also the treatment of depression in Coates & Swenson 2013).

These approaches aren’t patently absurd at first glance. Decision-making is, after all, a mental phenomenon and mental illness will certainly be related to it in some way—e.g. by affecting one’s preferences, perceptions, reasoning, and attention span. For instance, even if severe depression doesn’t significantly mitigate responsibility, it certainly affects one’s desires (e.g. a lack of interest in daily activities). So mental disorders may always affect responsibility to a small extent, whether by influencing one’s choices or one’s control over outcomes. But these sorts of factors can be relevant to some degree without mitigating blame, similar to one’s immediate environment, culture, genes, and mood. For example, while some emotions may hinder self-control, an adult is normally still fully blameworthy for attacking a child during a fit of rage. What interests many philosophers and policymakers is: When does a factor, like psychopathology, affect one’s capacities and abilities to a sufficiently high degree that it *mitigates* or *eliminates* responsibility?

It can certainly be tempting to think that mental disorders always or nearly always excuse. Responsibility and blame are the provinces of desert and punishment, and on one extreme psychopathology is thought of as a prime candidate for removing someone from the ordinary social practices in which these notions operate. The mentally ill are to be treated, not punished (or praised, presumably). Call this approach the *Naïve view*. On this view, an individual’s psychopathology implies something about their moral responsibility; mental illness as such affects individual responsibility.

Why might someone adopt the Naïve view? We aren't aware of any rigorous articulation or defense, but we can think of several possibilities.

First, one might argue that mental disorders are *pathological*, almost by definition, and so actions influenced by them inherit the property of being pathological or disordered. The idea here is that being classified as a disorder already tells us something about the possibility of an agent's being responsible for actions produced or significantly influenced by that disorder. However, just because an action is in some sense pathological, it doesn't imply that one is thereby excused. We'll see that there are clearly some cases in which one is still responsible for an action despite it being the result of a mental disorder.

A second rationale for the Naïve view is the idea that one isn't responsible for actions resulting from psychiatric disorders because one isn't *ultimately responsible* for having the disorder itself. The thought is reminiscent of "sourcehood" principles according to which, roughly, one is responsible for one's action only if one is responsible for what led to it (Strawson 1994). An initial problem for this rationale is that it might overgeneralize to ordinary actions that don't arise from the symptoms of a mental disorder. Ultimately, none of us are responsible for the sources of our actions if we trace the causal chain far enough back. So in order to make this kind of rationale applicable to mental illness in particular, as opposed to an uninformative general skepticism, one would have to argue that neurotypical people can be responsible for the springs of their actions in ways that those with mental disorders cannot. We doubt there is a clear sense of sourcehood such that all and only mental disorders will undermine it.

One can still appeal to sourcehood as part of a skeptical argument regarding the very possibility of responsibility. But such a skeptical thesis would establish the Naïve view for the wrong reasons. We're exploring whether there is a general relationship between moral responsibility and psychopathology that is grounded in facts internal to both of these concepts, not reliant upon a controversial position held for reasons that don't have to do with the specifics of mental illness.

Third, some might argue that mentally disordered individuals are simply *exempt* from our usual practices of holding people responsible. Like children and non-human animals, those with mental disorders are not appropriate targets of the "reactive attitudes," such as resentment, that support at least some of our core responsibility practices. Some have argued for this sort of conclusion for particular subsets of mental disorder, such as psychopathy (e.g. Shoemaker 2015). At least for some disorders with more pronounced and serious effects, a kind of "deep-rooted psychological abnormality" (Strawson 1962: 11) might warrant taking an "objective attitude" toward the individual, as a thing to be controlled or managed, rather than held responsible. Or perhaps the idea is that mental disorders lead to irresistible impulses that preclude proper control over one's conduct (cf. Wallace 1994: 169; contrast Pickard 2015). At any rate, whatever the prospects for exempting those with particular conditions, we'll see that such a conclusion is difficult to establish with respect to mental disorders generally or even an interesting subset of them.

Perhaps there are other plausible arguments for the Naïve view. We think it's already evident that the foregoing have several shortcomings. In fact, we'll argue that we shouldn't expect that there is any psychopathology for which possession of it implies anything definite about a person's responsibility for action. Ultimately, we think any informative version of the Naïve view, regardless of its rationale, will face the same general obstacle. What the view requires, and what we deny, is that the class of mental disorders (or some subset of them) is *unified* in a way that makes possible a generally supportable inference of the form: if an agent has a mental disorder, some conclusion about their moral responsibility follows.

If the Naïve view is not plausible, we must look more carefully at the distinctions among kinds of disorders, as well as the features of individuals most relevant to their responsibility. Indeed, our *Nuanced account* is decidedly unqualified, for we contend that there is no general relationship between moral responsibility and psychopathology; we must evaluate responsibility on a case-by-case basis. Alas, simplicity must give way to complication.

Other ethicists may seem to develop nuanced accounts but don't go far enough. David Shoemaker (2015), for example, has argued for an account that may seem similar to ours. He denies that mental disorders necessarily excuse, but thinks this is because the concept of responsibility itself is nuanced. He contends that there are senses of "responsibility" (such as "attributability" and "accountability"), only some of which apply to those with particular disorders, but which sense can correspond to which disorder varies. For example, he writes that "clinically depressed agents' attributability is (at least) mitigated" (143) and people with "high-functioning autism are (at least) impaired for accountability" (147). So, for particular kinds of responsibility, Shoemaker maintains what we're calling a version of "the Naïve view." Similarly, Carl Elliott (1996) doesn't contend that all mental disorders bear the same relationship to responsibility, yet he still argues roughly that psychopaths and those with compulsive disorders categorically aren't responsible, while people with personality disorders are.

Other ethicists have held views that look more nuanced. Nomy Arpaly, for example, writes that, "while many mental disorders do seem to provide exempting, excusing, or mitigating conditions, some do not, and with others things are complicated" (2005: 291; cf. also Feinberg 1970; Meynen 2010). However, even this statement suggests that some disorders categorically excuse. At any rate, we aim to contribute a more detailed argument and framework on which to base the Nuanced view.

### **3. How Theory Matters**

An initial complication is that some symptoms of mental disorders might affect responsibility but only on some theories. We'll see that there is some significant amount of agreement about which factors excuse. However, examining some points of contention helps to reveal the complex relationship between moral responsibility and psychopathology.

Unsurprisingly, there is an enormous philosophical literature on the necessary and sufficient conditions for responsibility. We cannot possibly canvas it all here. Instead, we propose to highlight a few features of agency that different theories of responsibility have emphasized, exploring the ways in which different disorders will be more or less threatening to those features. The aim is not to present a thoroughgoing typology of theories, but rather to examine some ways in which the answer to whether a mental disorder excuses may be theory-dependent.

We will focus on three features of agency that are central to discussions of moral responsibility: *choice*, *control*, and *coherence*. These features may overlap, and there is certainly room for disagreement as to which are necessary for moral responsibility. But, since our aim here is not to adjudicate between competing theories, we will speak fairly broadly and consider elements familiar to a wide variety of approaches to the questions of free will and responsibility. Additionally, we are skeptical that a single theory's preferred feature of agency can be privileged as *the* common sense thinking on the matter (cf. May 2014).

Some theories emphasize the notion of a free *choice* between genuine options. On such approaches, what responsibility requires is that the agent select from a range of available alternatives, electing and enacting one of them, unconstrained from external forces (Kane 1996). Accounts will vary, of course, as to the strength and stringency of these requirements. Nevertheless, these views often stress the agent's ability to consider a range of actions they might perform, and choose without constraint which one to execute. (While not required, these accounts tend toward accepting the incompatibility of responsibility and determinism. The thought is that determinism would preclude alternatives from which an individual could meaningfully choose.)

Other theories focus less on choice among alternatives and more on exercising effective *control*. On some views, it is appropriately recognizing and responding to the reasons one has (Fischer & Ravizza 1998). Other views favor characterizing control as a power to cause one's actions in a particular way (Clarke 1993). Uniting these views is the thought that the realm of the intentional is unique and uniquely important for responsible agency. Responsibility for what we do depends on our ability to identify and assess reasons, reasons upon which we can then act. These views also highlight how lack of control often undercuts one's responsibility. When an agent does something entirely by accident, it looks as though they are (to some extent) excused.

In contrast to the way in which an action or choice is brought about, some views of responsibility emphasize the *coherence* the action (and its motivations) has to the agent's psychology (Frankfurt 1971; Wolf 1987). Actions can reflect or express aspects of an agent's psychology to varying degrees and mental states can be more or less well-integrated into the agent's overall psychology. The guiding thought here is that an agent is more responsible the more reflective of their moral selves their action is or the better integrated its motivations. A deeply committed racist is all the more blameworthy for their racist remarks than one who makes an

out-of-character insensitive comment. Responsibility is undercut, therefore, when the action fails to manifest the agent's real commitments and values.

To see why the answer to whether psychopathology excuses may be theory-dependent, it will help to consider some disorders. Take obsessive-compulsive disorder (OCD), for example, which is characterized by persistent and unwanted thoughts or urges and repetitive actions performed as responses to the obsessions. Should someone who suffers from OCD be excused when, say, he knowingly breaks a promise to attend his daughter's piano recital in order to repeatedly wash his hands?

The answer seems to depend on which view we adopt. On views that stress the importance of choice, the fact that an action is a compulsion seems *prima facie* excusing. It isn't that the agent freely selected among their options. Instead, their options were unusually constrained, much in the way a bank teller has 'no choice' but to hand over the money at gunpoint (but see Pickard 2015). Similarly, on some control-based views, one may be excused if the nature of the compulsion undercuts the agent's ability to recognize or respond to the relevant reasons. The compulsive behavior quiets an intrusive urge, rather than reflects the agent's assessments of what was worth doing.

In contrast, a control-based view that only requires the agent to control the action intentionally by exercising special causal powers could still find the compulsive responsible, if the action was brought about in that way. Most strikingly, perhaps, on a coherence view, nothing internal to the view prohibits agents from endorsing or identifying with the obsessions that motivate their behaviors. For example, rather than seeing one's obsessive behaviors as alienating, one might endorse the life of trying to be meticulously clean or taking all safety precautions. It is thus possible for such agents to achieve the sort of integration required to be responsible, even if OCD will likely excuse on most versions of a coherence view. Indeed, OCD often varies in the degree to which patients have insight to their disorder. Those with more insight are presumably more aware of the irrationality of the obsessions, and so may be less likely to identify with those desires (cf. Shoemaker 2015: ch. 4).

Not all disorders will garner these same results. Consider major depressive disorder, partially characterized by depressed moods for most of the day, feelings of worthlessness, weight loss, and lowered motivation to engage in activities. On coherence views, such agents may be excused, for their motivations to remain indolent or shun interaction are not well integrated into their overall psychology. Nevertheless, depressed individuals may have no problems controlling their conduct, even when depressed. Their depression need not inhibit their ability to perform actions for reasons of which they are aware or to bring actions about via special causal relations. Whether or not a person with major depressive disorder is excused depends on the details.

More generally, theories premised on coherence must address whether a mental disorder expresses or masks one's true self. Yet, as Jeanette Kennett (2007)



points out, a patient's loved ones often say, "He's not really like this" or make similar proclamations. Many mental disorders (e.g. schizophrenia and dementia) set in well after one has an established personality and identity. Other disorders may be more difficult to dissociate from one's real self, especially those that tend to develop early in one's life, such as autism, Tourette's syndrome, and various antisocial personality disorders. Theories that do not rely on a real self, however, need not grapple so much with this issue. For example, someone with kleptomania may exhibit sufficient control over stealing some jewelry, whether or not it issued from her true self.

Just as we cannot canvas every approach to moral responsibility, we cannot consider here every disorder. We hope to have provided merely a sampling of the ways in which the symptoms of a psychiatric disorder might excuse, depending on the particular approach to free will or moral responsibility employed. To the extent that theories of responsibility will vary in the conditions or features of agency they prioritize, there will be variance in the degree to which the symptoms of particular disorders will be potential threats to responsibility. This complicated relationship between mental disorders and responsibility is enough to raise initial doubts about the Naïve view. No standard approach to moral responsibility supports a general inference from mental disorder to any particular claim about responsibility.

#### **4. Bypassed or Diminished Agency**

It's controversial whether moral responsibility requires choice, control, and coherence, but nearly all agree that it requires "agency." Agents believe things, intend things, and desire things. They make and revise plans, reflect on their motivations, and judge ends to be more or less worth pursuing. They seek to make their actions intelligible and conform to norms they accept. Regardless of whether all of these capacities are essential to agency as such, it seems only agents can be (morally) responsible for their actions. And, importantly, it seems that when individuals are responsible for an action it is due to features of their agency or its exercise.

A mental disorder plausibly excuses, then, if its symptoms yield an action that entirely *bypasses* one's agency, such that an outcome is the result of no action at all or an entirely unintentional one. Consider, for instance, narcolepsy, characterized by abnormal sleepiness during the day or lapses into sleep, and often accompanied by cataplexy, or sudden muscle weakness that often makes patients collapse. Suppose a narcoleptic is prone to sudden, unpredictable cataplexy, and during one such episode, drops the priceless vase he was carrying. Ordinarily, one might be blameworthy for dropping a vase. But it seems the mental disorder provides a ready, and full, excuse. His narcolepsy excuses him in a way that one who simply tripped or was distracted is not.

Now it might be argued that such excuses are limited. Once one becomes aware that one has a disorder, one has a responsibility to manage its effects. Suppose, for example, that a patient with schizophrenia responds well to treatment with an anti-psychotic medication like loxapine. If she discontinues use, however,

she typically has haunting hallucinations that cause her to violently attack those around her, misperceiving them as imminent threats. In such cases, where symptoms are expected and their effects can be mitigated, one might be responsible for harming others, even if it is due to hallucinations, because the patient knows such situations can be prevented by staying on the medication. In such cases, we might attribute responsibility for an outcome that resulted from diminished agential capacities by transferring the responsibility one has for not allowing those capacities to diminish in the first place (cf. the “transfer principle” in Summers & Sinnott-Armstrong 2015).

However, this oversimplifies the notion of tracing responsibility for acts to some prior opportunity to prevent them. Such scenarios are not common, as symptoms often first present without warning or can’t be managed anyhow. Moreover, even when one can knowingly manage symptoms, it remains controversial how best to justify transferring responsibility to outcomes from failures to take suitable precautions (cf. King 2014).

So, if narcolepsy can readily excuse, perhaps most other mental disorders do as well. However, we must be careful not to generalize too quickly. “Mental disorders” does not denote a homogenous class. Different disorders have different properties, they affect different capacities, and the ways they affect those capacities can vary widely. It would be surprising if, despite these differences, they all amounted to full excuses. Considering Tourette’s syndrome is instructive. Many people assume tics are entirely unintentional or involuntary and thus clear candidates for excusable actions not performed of the patient’s own free will. But patients with this syndrome often report acting voluntarily; it’s just increasingly difficult to overcome the impulse (Schroeder 2005). While those with Tourette’s syndrome may sometimes lack responsibility for their tics, this case illustrates that matters are often more complicated than they seem.

There is further reason to believe that we can’t generalize from cases like narcolepsy. The disorder seems to excuse dropping the vase so completely because the condition is *directly relevant* to the action in a way that does not hold true of every mental disorder. For example, contrast narcolepsy with someone suffering from OCD who experiences intrusive and unwanted thoughts that cause anxiety. In many cases, an episode of such intrusion will have no bearing on their ability to carry the vase. And so their mental disorder will not significantly affect their responsibility for dropping the vase, should they drop it. Thus, it seems that one’s disorder must not only compromise a capacity relevant to responsibility, but it also must be relevant to the act in question (Feinberg 1970: 273).

To further illustrate, suppose a man with attention-deficit/hyperactivity disorder (ADHD) assaults an individual at a sporting event after an angry confrontation that escalated from a disrespectful gesture into violence. ADHD involves difficulties in paying attention, staying focused, and organizing one’s life for success in, say, school or work. Given the nature of this disorder, it is unlikely that it played a crucial role in generating an aggressive action, especially one that might

typically arise in anyone without ADHD, as it simply involves relatively normal emotional reactions to a show of disrespect.

Now, ADHD can involve increased impulsivity, in which case the symptoms of the disorder may play a crucial role in an act of aggression. A similar analysis seems appropriate for conduct disorders affecting impulsivity, such as kleptomania, which can obviously play a key role in an act of stealing. Even if it's controversial whether kleptomania should excuse or mitigate responsibility, this is a case where the disorder quite clearly influences the act in question, such that someone in similar circumstances but without kleptomania likely would not steal. In the same vein, however, kleptomania would appear irrelevant to a case of acting out of aggression—indeed, irrelevant to most of the actions one performs.

Of course, it may seem uncharitable to say the Naïve approach holds that anyone with a psychiatric disorder is thereby excused from all of their actions, including those not influenced by the disorder's symptoms. However, we think that even a more charitable version of the view is problematic, for there are cases in which actions can be affected by a mental disorder without yielding mitigation or excuse.

Part of the reason psychopathology only sometimes excuses lies in the notion of capacity. By definition, mental disorders affect mental capacities, and some of these capacities are integral to freedom and responsibility. However, diminished capacity does not entail *lack* of capacity (Glannon 2011, ch. 3). Many disorders, such as autism, involve a varied spectrum. Some patients with autism may appropriately be described as having an *inability* to pick up on non-verbal social cues or even navigate the social world successfully whatsoever. Others, however, have only minor difficulties in this respect. Indeed, many disorders—from depression to bipolar disorder to psychopathy—are recognized to present along a continuum. The upshot is that, for many disorders, being diagnosed does not necessarily affect one's responsibility and certainly not in the same way as others in the diagnosed class. Sometimes the symptoms will be so slight so as to hardly diminish a relevant agential capacity.

Now, a proponent of the Naïve view might emphasize that mental disorders can excuse even when they don't entirely bypass one's agency; they just need to compromise it to a significant degree. Consider, for example, the famous case of *Clark vs. Arizona* (48 U.S. 735 2006). The defense argued that Clark did not kill a certain police officer intentionally because he thought the officer was a space alien, due to his paranoid schizophrenia (discussed in Morse 2011). Now the Supreme Court thought Clark still knew this was wrong, but other delusions or hallucinations might lead one to do something that, from one's own perspective, should be counted as morally permissible (cf. Broome et al 2010: 183). However, even in these cases, one's agency in the particular circumstances is substantially compromised.

It's certainly true that a mental disorder's symptoms can excuse wrongful acts without entirely bypassing one's agency. But normally the symptoms must significantly undercut one's agency. Consider, for example, ADHD which does not

appear to disrupt one's general capacity to form intentions and act on them. When someone with this disorder acts, their agency is not necessarily diminished substantially. One's ability to focus may be limited, but this may not distinguish ADHD from more ordinary conditions like being tired or upset. (Various mild forms of anxiety disorders may be similar in this respect.)

Further complexities loom, for there are many ways in which mental disorders might (or might not) significantly affect the exercise of one's agency. To illustrate, let's draw two cross-cutting distinctions that can be applied to categorize how disorders might present themselves.

First, a disorder can be episodic or static. *Episodic disorders* present themselves in (more or less) discrete instances. Narcolepsy, again, is a good example, as the associated loss of consciousness comes in discreet instances. Though a patient with narcolepsy is always possessed of their condition, it is only episodically activated, as it were. Similarly, while there is currently no cure for Alzheimer's, patients can within the same day go from experiencing relative clarity of mind to confusion, despite unchanged neurodegeneration (Palop et al. 2006). Other examples might include dissociative identity disorder, PTSD, bipolar disorder, specific delusions, and various phobias—each of which can manifest in discrete episodes, sometimes in response to specific triggers. In contrast, some psychopathologies are more *static*, like autism, depression, psychopathy, dementia, and generalized anxiety disorder. The ways in which each manifests is more likely to persist over time, with no clear boundaries. Perhaps their effects can wax and wane, but we wouldn't naturally carve them up into discrete episodes.

The second distinction concerns the degree to which a disorder impinges on one's agency. Narcolepsy, for example, has quite *global* effects: loss of consciousness undermines the affected agent's abilities across the board. In contrast, certain disorders may only be relevant to a subset of agential abilities, yielding more *local* effects. For instance, kleptomania, as a compulsion, presents as strong urges to steal, but it leaves other elements of an agent's psychology relatively untouched. Specific or "simple" phobias are another kind of example, as they are tied to certain cues, e.g. spiders, heights, or blood. Similarly, while some delusions can be systematic and elaborated (as in forms of schizophrenia), others are rather specific and relatively circumscribed, as when patients with Capgras take a familiar person to be an imposter (Bortolotti 2015).

These distinctions cross-cut, so either element of each can pair with either element of the other (see Table 1). Narcolepsy is an episodic, global condition, while severe schizophrenia may be a static condition, but still with global affects. Anxiety disorders may have localized effects, despite being static conditions, while some episodic conditions, like specific phobias, will similarly only present locally.

**Table 1: Kinds of Effects on Agency**

|               | <b>Episodic</b> | <b>Static</b>          |
|---------------|-----------------|------------------------|
| <b>Local</b>  | e.g. phobias    | e.g. anxiety disorders |
| <b>Global</b> | e.g. narcolepsy | e.g. schizophrenia     |

Drawing these distinctions further motivates the Nuanced Account. Regardless of whether a disorder has only local effects to agency or global ones, and irrespective of whether it operates episodically or statically, psychopathologies do not support a general effect on responsibility across all contexts. The case in question may not be a relevant episode, or it might not involve the agential capacities the disorder affects. Even global, static disorders, which might seem to lend the most support to the Naïve view, don't support the general inference. Having a disorder with widespread effects on agential capacities that operates across time doesn't guarantee that those effects were present in the given case. In short, as the Nuanced view holds, establishing that an agent has a particular psychopathology tells us nothing about their responsibility.

## **5. When Symptoms Enhance**

Let's consider one last reason to favor the Nuanced account. The Naïve view says that, in general, mental illness excuses relevant behavior or significantly mitigates blame (and praise). This is further suspect given that some symptoms are capable of *enhancing* one's responsibility.

We often associate mental disorders with deficits in various capacities. However, what might normally disrupt an ordinary person's life may turn out to be beneficial in certain circumstances. The early effects of dementia, for example, might free one from haunting memories of abuse (Earp et al. 2014). Similarly, while typically unpleasant, anxiety can alert one to relevant information and reduce uncertainty about potential threats to oneself or others (Kurth in press). Even delusions can serve as defense mechanisms to help patients overcome obstacles, such as psychological trauma or low self-esteem, by making sense of their experiences (Bortolotti 2015), not unlike the rationalizations frequently found among neurotypical individuals (cf. Summers 2017; May forthcoming).

So it's possible that, even if harmful overall, a mental disorder's symptoms may enhance certain capacities, such that, surprisingly, one becomes, as Stephen Morse (2006) puts it, "hyper-responsible." Here we may find parallels with the enhanced responsibilities of those possessing advanced skills or knowledge (Vincent 2013), as when we hold only physicians accountable for failing to provide medical advice for someone ill on an airplane. Perhaps enhanced capabilities merely increase the number or kinds of responsibilities, rather than degrees of responsibility itself (Glannon 2011, p. 120). Nevertheless, enhanced responsibility can result from an increase in mental capacity, whether from learning, ingesting a pill, or mental disorder.

Conditions that increase attention are natural candidates. Consider, for example, disorders involving episodes of hypomania, such as bipolar disorder (cf. Arpaly 2005: 290; Morse 2006; Vincent 2013). Unlike mania, episodes of hypomania don't severely disrupt one's life, but in either case patients often need much less sleep (e.g. "feels rested after only 3 hours of sleep"), have a "flight of ideas," and increased "goal-directed activity" (DSM-5, 124). Some of these symptoms read as desired effects of cognitive enhancers, like amphetamine or modafinil, even if they can be rather disruptive in many contexts or when combined with other typical symptoms (e.g. engaging in risky behavior). However, it is important to note that patients often don't have knowledge of, or control over, when they will have any purportedly enhanced capacities during hypomania, and thus it may seem inappropriate to hold them more accountable (Turner 2010), at least in many circumstances.

Perhaps this is less of a limitation for another candidate for enhancing responsibility: OCD. The obsessive thoughts are often tied to a specific anxiety or source of distress, such as uncleanliness (e.g. excessively washing hands) or danger (e.g. locking a door repeatedly). Those with scrupulosity, in particular, are especially concerned to behave morally, often concerning others believed to be in need (Summers & Sinnott-Armstrong 2015). Such hyper-awareness and sensitivity to morally relevant considerations may, in certain circumstances, enhance capacities relevant to freedom and responsibility. For example, suppose Saul leaves the door unlocked and a family of raccoons ransack the house. We might hold Saul more responsible given his hyper-awareness of whether the door is locked. Similarly, while we might praise Sally for mindlessly locking the office door and preventing a robbery, we'd likely praise Saul even more if it resulted from his heightened concern for safety. But notice that it's not the possession of psychopathology as such that enhances one's responsibility; it's the specifics of a certain symptom and circumstance.

## **6. The Need for Nuance**

The diversity of ways in which the symptoms of mental disorders affect action makes them an extremely heterogeneous class, such that there is no supported general inference from having a mental disorder to any claims about one's moral responsibility. Mental illness can sometimes excuse actions, provided the symptoms significantly undermine some feature of one's agency that generated the act in question. While it remains controversial which features of agency are necessary for responsibility (e.g. choice, control, or coherence), there is substantial overlap among the various theories of responsibility. Even in this region of overlap, however, it's clear that mental illness sometimes excuses and sometimes doesn't. Many disorders have local, rather than global, effects on specific agential capacities, sometimes in fairly discrete episodes that leave the rest of one's life relatively isolated from the disorder's effects. Moreover, many disorders lie on a spectrum, and at many points one's agential capacities are diminished to some degree but not sufficiently diminished.

We thus ought to reject the Naïve view and adopt a Nuanced account of the relationship between moral responsibility and psychopathology. There is no reason to believe that having a mental disorder generally makes one less responsible than those who enjoy better mental health.

To further illustrate, let's briefly return to the legal cases with which we began. Since these are actual events, they are complex and we certainly lack sufficient evidence to know all of the relevant facts. Nevertheless, we can try to apply some of the lessons developed here. In both cases it is plausible that symptoms from the defendants' mental disorders did contribute to the criminal acts in question. In Yates's case, her psychiatrist warned about having a fifth child when Yates attempted to commit suicide soon after having her fourth. So it seems her major postpartum depression and various delusions plausibly contributed to her intention to end her children's lives just months after the birth of her fifth. In Routh's case, his PTSD plausibly contributed to the murders, as his attack was on two military veterans at a shooting range. However, arguably these disorders do not typically yield actions that bypass one's agency entirely. Thus, as far as morality is concerned, we may unfortunately be left with theory-dependent considerations—e.g. about whether these defendants lacked sufficient choice, coherence, or control. Arguably, these are precisely the sorts of considerations to which the law should be sensitive, even if current practice has yet to be suitably responsive.

Indeed, despite public discourse on such high-profile cases, the law generally operates with something like our Nuanced view, applied specifically to legal liability. Lawyers can't simply establish that their clients have a mental disorder. They must show that the relevant symptoms of the disorder causally contributed to the act in question and compromised some psychological capacity relevant to the elements of legal liability, such as negating the requisite *mens rea* or voluntary act.

A similar approach arises in medical settings. The competence of patients to make an informative and autonomous decision is not determined simply by categorizing them as having a mental disorder. Rather, doctors and caregivers evaluate the specific case at hand, quite independently of how the patient fits into categories found in the DSM (see e.g. Appelbaum 2007).

Our view may appear harsh on the victims of mental illness, suggesting that there are many circumstances in which they are responsible, even for acts substantially influenced by their symptoms. However, the Nuanced account precisely shirks any such general claims about moral responsibility and psychopathology. Sometimes one is responsible despite a mental illness, sometimes not.

Indeed, we resist the idea that those with psychiatric disorders are *always* appropriate targets of praise and blame. Consider, for example, Pamela Bjorklund's (2004) view that those with personality or character disorders are "rightly 'responsible' for their actions" (190). Her position is partly a defense of therapies that encourage such patients to "take responsibility" for their behavior, including their aim of achieving mental health. More generally, Hannah Pickard argues that

“psychopathology does not strip people of free will” (137), at least not by eliminating choice or the ability to do otherwise. Of course, if read as mere denials of the idea that having a mental disorder always mitigates one’s blame, then we agree. But we caution against any generalizations of the form that possession of a mental illness implies anything about responsibility for action.

Nevertheless, the more nuanced picture developed here does allow us to treat many who suffer from psychopathologies as autonomous agents whose decisions about how to live we should respect. Indeed, being responsible allows for the possibility of praise and discourages a passive attitude toward one’s condition that can impede improvement (Pickard 2015). Our approach also points to ways in which those with mental disorders may exist on a spectrum with the neurotypical individual, who nonetheless may be exhausted, hungry, or prone to distraction, recalcitrance, or fear. Thus, the Nuanced account might be viewed as supporting a neurodiversity or integrationist approach to mental illness—where those with psychopathologies and the neurotypical are sometimes more alike than they are different—which may help to reduce the stigma of mental illness.

**Acknowledgements:** For helpful feedback on earlier drafts, we thank: Justin Caouette, Justin Clarke-Doane, Neil Levy, Kathryn Tabb, and Jesse Summers. A version of this paper was presented at a graduate seminar at Columbia University. Many thanks to Kathryn, Justin, and their thoughtful students for a profitable discussion that greatly shaped the final version of the paper.



## References

- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC.
- Appelbaum, P. S. (2007). Assessment of patients' competence to consent to treatment. *New England Journal of Medicine*, 357(18), 1834-1840.
- Arpaly, N. (2005). How it is not "just like diabetes": Mental disorders and the moral psychologist. *Philosophical Issues* 15 (1):282-298.
- Bjorklund, P. (2004) 'There but for the grace of God': moral responsibility and mental illness. *Nursing Philosophy* 5: 188-200.
- Bortolotti, L. (2015). The epistemic innocence of motivated delusions. *Consciousness and Cognition* 33(C): 490-499.
- Broome, M. R., Bortolotti, L., & Marnett, M. (2010). Moral responsibility and mental illness: A case study. *Cambridge Quarterly of Healthcare Ethics*, 19(02), 179-187.
- Clarke, R. (1993). Toward a Credible Agent-Causal Account of Free Will. *Noûs* 27(2): 191-203.
- Coates, D. J., & Swenson, P. (2013). Reasons-responsiveness and degrees of responsibility. *Philosophical Studies* 165(2): 629-645.
- Dart, T. (2015). "'American Sniper' jurors: defendant was using PTSD diagnosis as an excuse." *The Guardian*. Retrieved from <http://www.theguardian.com/us-news/2015/feb/25/american-sniper-trial-jury-eddie-ray-routh-ptsd-excuse>.
- Earp, B. D., Sandberg, A., Kahane, G., & Savulescu, J. (2014). When is diminishment a form of enhancement? Rethinking the enhancement debate in biomedical ethics. *Frontiers in Systems Neuroscience* 8(12): 1-8.
- Edwards, C. (2009). Ethical decisions in the classification of mental conditions as mental illness. *Philosophy, Psychiatry, & Psychology*, 16(1), 73-90.
- Elliott, C. 1996. *The Rules of Insanity: Moral Responsibility and the Mentally Ill Offender*. Albany, NY: SUNY Press.
- Feinberg, J. (1970). What is so Special About Mental Illness? *Doing and Deserving*. Princeton: Princeton University Press, 272-92.
- Glannon, W. (2011). *Brain, Body, and Mind: Neuroethics with a Human Face*. Oxford University Press.
- Kennett, J. (2007). Mental disorder, moral agency, and the self. In Bonnie Steinbock (ed.), *The Oxford Handbook of Bioethics*. Oxford University Press. 90-113.
- King, M. (2014). Traction without Tracing: A (Partial) Solution for Control-Based Accounts of Moral Responsibility. *European Journal of Philosophy* 22 (3): 463-482.
- Kurth, C. (in press). *The Anxious Mind*. Cambridge, MA: MIT Press.
- Levy, D. A. (2003). Neural Holism and Free Will. *Philosophical Psychology* 16(2): 205-228.
- Levy, N. (Ed.). (2013). *Addiction and self-control: Perspectives from philosophy, psychology, and neuroscience*. Oxford University Press.
- Maibom, H. L. (2008). The Mad, the Bad, and the Psychopath. *Neuroethics*, 1(3), 167-184.
- May, J. (2014). On the Very Concept of Free Will. *Synthese* 191 (12):2849-2866.
- May, J. (forthcoming). *Regard for Reason in the Moral Mind*. Oxford University Press.
- Meynen, G. (2010). Free will and mental disorder: Exploring the relationship. *Theoretical Medicine and Bioethics* 31 (6):429-443.
- Morse, S. J. (2006). "Moral and Legal Responsibility and the New Neuroscience." In Judy Illes (ed), *Neuroethics: Defining the Issues in Theory, Practice, and Policy*. Oxford: Oxford University Press, 33-50.
- Morse, S. J. (2011). "Mental Disorder and Criminal Law." *The Journal of Criminal Law and Criminology* 101(3): 885-968.

- Newman, M. (2006). "Yates Found Not Guilty by Reason of Insanity." *New York Times*. Retrieved from <http://www.nytimes.com/2006/07/26/us/26cnd-yates.html>.
- Palop, J. J., Chin, J., & Mucke, L. (2006). A network dysfunction perspective on neurodegenerative diseases. *Nature* 443(7113): 768–773.
- Pickard, H. (2015). Psychopathology and the Ability to Do Otherwise. *Philosophy and Phenomenological Research*, 90(1), 135–163.
- Schroeder, T. (2005). Moral responsibility and tourette syndrome. *Philosophy and Phenomenological Research* 71 (1):106–123.
- Shoemaker, D. (2015). *Responsibility From the Margins*. New York: Oxford University Press.
- Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies* 75(1–2): 5–24.
- Summers, J. S. & Sinnott-Armstrong, W. (2015). Scrupulous agents. *Philosophical Psychology* 28(7): 947–966.
- Summers, J. S. (2017). Post hoc ergo propter hoc: some benefits of rationalization. *Philosophical Explorations* 20(sup1): 21–36.
- Tabb, K. (2015). Psychiatric Progress and the Assumption of Diagnostic Discrimination. *Philosophy of Science*, 82(5), 1047–1058.
- Turner, A. J. (2010). Are disorders sufficient for reduced responsibility? *Neuroethics* 3 (2):151–160.
- Vincent, N. (2013). "Enhancing Responsibility." In N. Vincent (ed.), *Neuroscience and Legal Responsibility*. Oxford University Press, pp. 305–33.
- Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Harvard University Press.
- Wolf, S. (1987). "Sanity and the Metaphysics of Responsibility." In F. D. Schoeman (ed.), *Responsibility, Character, and the Emotions*. New York: Cambridge University Press, 46–62.