

## Kant on Mental Disorder II: Philosophical Implications of Kant's Account DRAFT

In a discussion of the possibility of a priori knowledge, Russell critiques Kant's account of such knowledge on the grounds that, among other things, Kant's account depends upon contingent facts about human nature. As Russell explains,

The thing to be accounted for is our certainty that the facts must always conform to logic and arithmetic. To say that logic and arithmetic are contributed by us does not account for this. Our nature is as much a fact of the existing world as anything, and there can be no certainty that it will remain constant. It might happen, if Kant is right, that to-morrow our nature would so change as to make two and two become five. This possibility seems never to have occurred to him.<sup>1</sup>

Russell's problem arises because Kant seems to ground the a priori necessity of truths of mathematics (and even logic) on the conditions of possibility of *our* sensing and thinking about the world. Because *we* (human beings) perceive the world in Euclidian space-time and think about it using various logical categories, truths of geometry, arithmetic, and logic must be true in any world that is to be a world for us. But, Russell suggests, that makes these truths ultimately contingent upon human nature, and a truth that is contingent upon human nature cannot be the sort of a priori – and hence necessary – truth that Kant sought in his *Critique of Pure Reason*.

Russell rightly points out that the possibility of radically different cognitive structures is a serious potential problem for Kant, but he wrongly claims that this possibility did not occur to Kant. Kant himself often points out that one cannot explain why human cognitive faculties are the way that they are. And Kant is well known (presumably even to Russell) to have discussed *non-human* cognitive variation, ultimately insisting that “we cannot decide” whether non-human rational beings have the same cognitive structure as human beings (B72). But, in texts that are much less well known (almost certainly not studied by Russell), Kant explicitly discusses the possibility that even human beings might have radically different cognitive structures. He calls such states “craziness” (or “insanity,” *Abernwitz*; *vesania*) and explains that “in this . . . kind of mental derangement there is not merely disorder and deviation from the rule of the use of reason, but also *positive unreason*; that is, *another* rule, a totally different standpoint into which the soul is transferred, so to speak, and from which it sees all objects differently,” such that the insane person “imagin[es] that he conceives the inconceivable” such as “the squaring of the circle” (7:216). In the insane, Kant finds precisely the sort of contingency of human nature that makes even the a priori truths of mathematics become, at least to the afflicted person, untrue.

Mental disorder potentially raises serious problems for Kant in several key junctures of his philosophical project. With respect to Kant's metaphysics and epistemology, the possibility of “a totally different standpoint” (7:216) seems to compromise the transcendental structure of Kant's arguments for a priori synthetic truths. Equally important, the possibility of mental disorder might seem to introduce into the behavior of at least some human beings a sort of determinism that would compromise the applicability of Kant's moral philosophy, which is highly dependent upon rational autonomy and a very strong conception of human freedom. Strikingly, however, while he has a conception of

---

<sup>1</sup> Bertrand Russell, *The Problems in Philosophy*, London: Oxford University Press, 1912, reprinted by Prometheus Books, 1998, p. 87. vanCleve 1999: 37-41 has a helpful discussion of this problem.

mental disorder that could (and perhaps should) have been discussed in the context of his theoretical or practical philosophy, Kant in fact reserves his treatments of mental disorder almost entirely to a different and long neglected arena of his thought: pragmatic anthropology.

In a previous article, I have offered a detailed overview of Kant's account of mental disorder and the way in which that account fits into his pragmatic anthropology more generally. For the purpose of this article, my goal is three fold. First, I aim to show why Kant offered a theory of mental disorder. Given the potential problems that such an account raises for his Critical philosophy more generally, one might expect Kant to either ignore mental disorders or bring them up specifically to address concerns like Russell's. But while Kant does not ignore mental disorders as such, he also does not introduce them in order to address the kinds of philosophical concerns raised by critics like Russell. Thus the first part of this article will be focused on laying out the purposes that *Kant* primarily had in mind in introducing his theory of mental disorder. Second, I turn to philosophical problems that Kant's theory of mental disorder seems to raise in the context of his Critical philosophy. These problems include the metaphysical and epistemological ones raised by Russell as well as problems for his moral philosophy that arise from Kant's discussions of disorders that affect volition. While Kant has some specific comments addressing the latter, he says virtually nothing explicit regarding the former sorts of problems. Nonetheless, Kant's theory of mental disorders *can* be well integrated into his Critical philosophy and doing so reveals important new aspects of Kant's approach to both cognition and volition. Finally, I turn to a brief conclusion in which I consider the implications of Kant's approach to mental disorder for contemporary psychology and psychiatry.

Before turning to these three main goals of this article, however, I offer a *very* brief recapitulation of Kant's theory of mental disorder.

### **I. Kant's Theory of Mental Disorder: A Very Brief Overview**

As I have laid out in detail elsewhere, Kant's theory of mental disorder fundamentally has a tripartite structure, corresponding to his tripartite psychology. For Kant, mental disorders can affect each of the three fundamental human psychological faculties: cognition, feeling, and desire. The faculty of cognition is itself sub-divided into a variety of different cognitive powers – the senses, the imagination, judgment, the understanding, and reason – and cognitive mental disorders can be distinguished by which cognitive power is affected and how. For example, dementia (*Wahnsinn*) is a deranged imagination, while craziness (*Aberwitz*) is a deranged reason. In addition, Kant distinguishes between cognitive disorders that are mere deficiencies in a cognitive faculty and those that are positive forms of derangement. Thus stupidity (is a deficiency of judgment whereby one simply lacks the ability to figure out whether a particular case falls under a general rule, while insanity (*Wahmwitz*) is a derangement of judgment whereby one groups together disparate particulars under false universals. Further, Kant adds melancholia and hypochondria as cognitive disorders distinct from those that fall under more general groupings. With respect to the disorders of feeling and desire, Kant treats all disorders of feeling under the general name of “affects” and disorders of desire under the general name of “passions.” Both involve a state wherein a particular feeling or desire overpowers the reflection needed to compare that feeling or desire with others, and thus a single feeling or desire motivates action without (sufficient) reflection. Finally, Kant includes accounts of the origins of various mental disorders and ways of treating them.

## II. Why does Kant provide a theory of mental disorder?

On the one hand, Kant includes his account of mental disorder in his pragmatic anthropology rather than in his Critical works, and consequently does not focus in his account on the philosophical implications of mental disorders for his Critical philosophy. On the other hand, Kant was not a physician, and Kant's theory of mental disorder is not a clinical theory in that it is not primarily meant to be applied by physicians seeking to cure mental disorder. Kant (in)famously<sup>2</sup> claims, in fact, that for determining whether or not a person is "mad" (*verrückt*), one must refer "not . . . to the medical faculty but . . . to the philosophical" (7: 213). And Kant specifically excludes those forms of "raving" that count as physical illness from the category of madness, or derangement (7: 213). So if it has neither a clear philosophical purpose nor a clear medical purpose, what precisely is the point of Kant's "anthropological" discussion of mental disorder?<sup>3</sup>

### *a) Clinical Uses: Prevention and cure of mental disorder*

First, one might expect Kant's account to focus on the prevention and cure of mental disorders. Not only is this a common reason for diagnosing mental disorders today, but it seems to fit extremely well with Kant's explanation of what *pragmatic* anthropology involves (7:119), as well as with Kant's explicit reason for distinguishing affects and passions (7: 251). As discussed in detail elsewhere<sup>4</sup>, Kant does provide advice on the prevention and, to a lesser extent, the cure of mental disorder. For example, Kant suggests that the reading of novels, by making distraction habitual, can become a cause of dementia, a serious form of mental derangement (see 7:207), so that avoiding novels is a way to help ward off dementia. And he proposes art and culture as ways of refining feeling and thereby treating affects (25:622-23). But there is considerably less detail regarding prevention and cure than one might expect in a discussion of mental disorder, especially one that is meant to be pragmatic. Several mental disorders are discussed without any mention of preventative or palliative measures; and some, such as dementia or insanity, are explicitly said to be incurable (7:215). Moreover, even when Kant does suggest preventative measures or cures, he does not offer a systematic account of these, instead merely dropping occasional suggestions.

To some extent, the limited nature of Kant's account here is actually forced by his insistence upon a pragmatic anthropology that is at the same time *not* physiological. Kant defines "pragmatic" anthropology by contrast with two features that he perceived to be intimately related in Platner's *Anthropology*: a purely theoretical emphasis and a focus on the physical (neurological) bases of mental states (7: 119). Because Kant perceived Platner's physiological emphasis as a contributing factor to the inability of his theory to yield practical results, Kant sought a more philosophical and *thereby* more pragmatic approach to medicine. To some extent, this pessimism about physiological approaches to mental health may be due to Kant's own experience with psychiatric medicine. In the same letter in which Kant explains to his friend and former student Marcus Herz that he will avoid "subtle and, to my view, eternally futile inquiries as to the manner in which bodily organs are connected with

---

<sup>2</sup> See Foucault, 2006, 123, 126.

<sup>3</sup> For the purpose of this section, I ignore the purely personal reasons that Kant has for his theory of mental disorder. Much of Kant's account of mental disorder can be understood in the context of his coming to terms with his own propensity to hypochondria and the particular awareness that this gave Kant of the fragility (but also the potential for self-control) of the human mind. For more on this theme, see (Frierson xxx; Shell, 1996, 268-305).

<sup>4</sup> Frierson xxx

thought,” he also reports autobiographically his bad luck with medicine to treat his own ailments:

I now feel much better than before. The reason is that I now understand much better what makes me ill. Because of my sensitive nerves, all medicines are without exception poison for me . . . . Study the great variety of constitutions. My own would be destroyed by any doctor who is not a philosopher. (10: 144)

At least in part because of his personal experience, then, Kant was extremely pessimistic about physiological-medicinal cures for mental disorder. The range of cures that he is able to offer for mental disorder is limited to those within the repertoire of a purely philosophical “physician of souls” (7:251). And thus Kant’s pragmatic anthropology ends up, for at those mental disorders that are at base physiological, being deeply *un*pragmatic.

Still, Kant’s approach to the prevention and cure of mental disorder, precisely because it is a sort of *Weltweisheit* (philosophy of worldly wisdom) rather than a branch of medical science, provides a kind of practical advice for dealing with the danger of mental disorder in one’s own life. In the same sort of way that Kant endorses autonomy in his ethics and “emergence from one’s self-incurred immaturity” in his essay “What is Enlightenment,” he offers an account of the “power of the mind to master its morbid feelings by sheer resolution” (7:97). This power has limits, and it will not apply to all (and not to the most severe) forms of mental disorder. But insofar as Kant provides advice for dealing with mental disorder, this advice is generally directed towards steps that one can take to prevent or cope with mental disorder *oneself*.

There is a closely related purpose of Kant’s taxonomies of mental disorder, one more in keeping with the generally practical character of Kant’s anthropology than a clinical science would have been. Kant’s anthropology is not only designed to help individuals understand themselves. Kant also crucially sees his anthropology as providing the tools for effective interactions with others, and especially with *diverse* others. He explains in his discussion of different national characters, for example,

In an anthropology from a pragmatic point of view . . . the only thing that matters to us is to present the character of each, as they are now, in some examples, and, as far as possible, systematically; which makes it possible to judge what each can expect from the other and how each could use the other to his own advantage. (7:312)

In the context, Kant is focused on giving the characters of people of different nationalities, but the same can be said about mental disorders. By showing the differences between different mental disorders, Kant can help his students and readers interact better with mentally ill people, even if their disorders can be neither prevented nor cured.

We must trouble ourselves to form the way of thinking and the capacities of those people with whom we have to do, so that we are not too hard nor too offensive. So we are taught anthropology, which shows us how we can use people to our ends. (25:1436)

Kant’s discussion of the ways that one without judgment can nonetheless be productive in matters of understanding (7:204), or how those with dementia, though incurable, are also harmless (7:214), or how to calm one in danger of rage (7:252) all help one avoid being excessively “offensive” while still protecting oneself and even furthering one’s interests.

#### *b) Legal and moral purposes*

Given Kant’s prominence as a moral philosopher, one might expect Kant’s theory of mental disorder to address the issue – increasingly important today – of legal implications of mental disorder. And Kant does provide at least some discussion of these juridical issues. In *Anthropology*, Kant considers the case where “someone has intentionally caused an

accident, [and] the question arises whether he is liable and to what extent; consequently, the first thing that must be determined is whether or not he was mad at the time” (7:213). Kant here insists, “In this case the court cannot refer him to the medical faculty but must refer him to the philosophical faculty (on account of the incompetence of the court)” (7:213). The reason for this is not that the issue of madness is a purely ethical one or one calling for specifically philosophical expertise, but rather because, for Kant, non-physiological empirical *psychology* is still a part of philosophy: “the question of whether the accused at the time of his act was in possession of his natural faculties of understanding and judgment is a *wholly psychological question*” (7:213, emphasis added). What Kant means to exclude from the juridical context is *physiology*. Moreover, the *reason* that Kant aims to exclude physiology is not that it cannot rule on these matters in principle, but only because the physiological bases of mental phenomena are not sufficiently well understood:

[A]lthough a physical oddity of the soul’s organs might indeed sometimes be the cause of an unnatural transgression of the law of duty (which is present in every human being), physicians and physiologists in general are still not advanced enough to see deeply into the mechanical element in the human being so that they could explain, in terms of it, the attack that led to the atrocity, or foresee it (without dissecting the body). (7:213-4)

For Kant, the fundamental matter at issue in juridical contexts is one’s mental state. And in the 18<sup>th</sup> century physiology with which Kant was familiar, medical science was insufficiently advanced to make determinations of physiological causes of criminal behaviors. Thus philosophers are better suited than doctors to diagnose the sorts of cognitive failings that are exculpatory.

For Kant, the key to legal imputation is the possession of normally functioning (though not necessarily flawlessly functioning) higher cognitive faculties.<sup>5</sup> In his lectures on ethics, Kant links these faculties to action through freedom: “we can attribute a thing to someone, yet not impute it to him; the actions, for example, of a madman or drunkard can be attributed though not imputed, to them. In imputation the action must spring from freedom” (27:288). Here it might seem as though the higher faculty of desire rather than the higher cognitive faculties should be most relevant. Generally, however, Kant takes the presence of higher cognitive faculties to be indicative of freedom in the higher faculty of desire. (The one exception would be action due to affects, which presumably would be non-imputable.) Thus insofar as one’s action is due to illnesses of the cognitive faculty, one can be excused from legal responsibility.

Despite offering at least some detail here, however, Kant seems to want to reserve judgment about imputation to philosophers as a way of avoiding an overuse of mental disorder as an excuse for misdeeds. In *Anthropology*, Kant’s only example of mental disorder as an excusing factor in court is an example of the *misuse* of a classification of mental disorder:

[I]n the case of a woman who killed a child out of despair because she had been sentenced to the penitentiary, such a judge declared her insane and therefore exempt from the death penalty. – For, he said, he who draws true conclusions from false premises is insane. Now this woman adopted the principle that confinement in the penitentiary is an indelible disgrace, worse than death (which is quite false), and came to the conclusion, by inference from it, that she deserved death. – As a result she was

---

<sup>5</sup> For more on the relationship between these higher faculties and moral and legal responsibility, see Patrick Frierson, forthcoming.

insane and, as such, exempted from the death penalty. – On the basis of this argument it might easily be possible to declare all criminals insane, people whom we should pity and cure, but not punish. (7: 214n)

For Kant, the judge was simply operating with an excessively naïve psychological picture of mental disorder. Kant seems to consider mental disorder a possible excusing factor in juridical contexts, but his primary concern is with how it can be overused there.

Although Kant downplays the role of mental – and especially cognitive – disorder in determining *juridical* imputability, he does highlight, in his *Metaphysics of Morals*, important implications of the difference between affects and passions for *moral* accountability. In particular, Kant insists that although affects and passions both hinder morality and both interfere with normal processes of rational choice, affects indicate merely a “*lack of virtue*” while passions are “*properly evil*” (6:408). Kant justifies this important distinction based on the specific nature of each mental disorder and especially based on the faculty of soul that each affects. Affects, which precipitously intensify one’s faculty of feeling to the point that reflection, and thereby choice, is precluded, indicate merely a “weakness . . . which can indeed coexist with the best will” (6:408). One might murder someone out of rage, but one’s power of choice is not involved. One might fail to provide easily provided assistance to someone in desperate need, but only because one is literally incapable of formulating the relevant desires. Acting or failing to act due to affects is morally similar to acting or failing to act when one is asleep. In both cases one is morally unconscious, so to speak, rather than morally corrupt. Thus affects hinder the development of virtue only in the negative sense. One is not made *less* virtuous because of them, but one does not progress in virtue. Unlike affects, passions operate *through* the power of choice, through choosing to pursue the object of one inclination without reflection upon other possible objects of choice. As a result, “passions are not, like affects, merely *unfortunate* states of mind full of many ills, but are without exception *evil* as well” (7:267). Passions do not preclude choice but pervert it. One with a passion *deliberately* rejects all interests, including the moral law, in favor of the pursuit of an inclination. So all passions are utterly incompatible with having a good will. They are “*properly evil*.”

As important as this distinction between affects and passions is (and I will discuss its philosophical implications in even more detail in the next section), Kant *uses* his discussions of affects and passions for an even more important purpose: to highlight the crucial role that character as a form of self-mastery plays in his moral anthropology. In the published *Anthropology*, and even more prominently in his lectures on anthropology (see especially 25: 425-6, 623-4, 1526-30), Kant uses his account of affects and passions to introduce his conception of *character*, which for Kant is that “property of the will by which the subject binds himself to definite practical principles” (7: 292). In his lectures, Kant makes this connection explicit. In one set of lectures, Kant explains that in order to prevent the outbreak of affects and passions, “one must so behave that one sets oneself to act according to principles” (25:588). And Kant ends the first half of his anthropology in his 1775-76 lectures with “general remarks on the passions and affects” (25:613). He concludes this section,

All these [passions and affects] are set against the steadfastness, wherein one does not deviate from his principle and persists firmly in his decision. It is important to have firmness in one’s decisions and not to depart from one’s principle, and better to endure disadvantage than to let go one’s principle. When then a human being knows

something for sure, he carries it out. But whoever is not steadfast therein often seizes a principle that he is certain that nothing will come of, because he knows that already he has frequently broken principles. Thereby the human being is a windbag in his own eyes, he believes himself capable of nothing more, from which springs hopelessness. That is a comfortless condition, when one always postpones hope. So it is with late conversions. So it is with other things, that one wants to break the habit of, e.g., oversleeping, for it always calls, just one more time but then no more, and so one philosophizes himself further away from carrying out [his resolution]. In such a condition one never has hope to become better; this is an important point in morals. One must therefore seek to hold himself ever so promptly to this word as to others. From that springs a firm confidence in ourselves. One who knows to make himself such that he can live with himself in peace is steadfast. (25:624)

With this passage, Kant concludes his discussion of two important forms of mental disorder, but he also introduces the much more general topic of acting in accordance with principles. And this sort of action in accordance with principles is what Kant calls “character,” and it is arguably the most important connection between Kant’s pragmatic anthropology and his moral philosophy. For Kant, in fact, the establishment of character is one of the most difficult obstacles to the establishment of *good* character.<sup>6</sup> Although Kant admits the possibility that one might act according to principles that are evil, he also claims that the majority of people who fail to live lives consistently in accord with the moral law fail to do so in large part because they fail to live consistent lives at all.

Now in general, lack of character is not mental disorder. But, first, lack of character can *lead* to passions and affects. And, second and more importantly, any lack of character bears an important *similarity* to mental disorder. Affects and passions both involve action without full deliberative self-control. Similarly, Kant aims to show, action even with full deliberation can fail to truly manifest self-control insofar as this deliberation is fleeting. Only by firming rooting oneself in consistent and principled action can one truly govern one’s own life. The failure to do this does not in itself constitute mental disorder, but it is a failure that is similar to and even leads to mental disorder. The ultimate effect of Kant’s transition from the mental illnesses of passions and affects to a general exhortation to (moral) character, then, is that ordinary human failings in acting consistently in accordance with principles are shown to be both a precursor to and a mild version of mental illness. In that sense, Kant’s discussion of affects and passions provides an entertaining way of highlighting some unusual failings of will that are at the same time paradigmatic for the ways in which all human beings fail to live up to the moral demands of principled action. Mental illness becomes, in this case, an illustrative special case of moral disorder. And the discussion of how to prevent (these) mental illnesses becomes an introduction to principles for a fully healthy and ultimately moral life.

### *c) Religious purposes: superstition, enthusiasm, and mental disorder*

These directly moral and pragmatic considerations are not the only ways in which Kant’s theory of mental disorder intersects with his broader moral and social concerns. One of the most important reasons that Kant spends so much time discussing mental disorders is as a part of his lifelong attempt to develop a “religion within the boundaries of mere reason.” Kant’s first published work on mental disorder – “On the Maladies of the Mind” (1764) – was written, at least in part, in response to general public interest in a Polish

---

<sup>6</sup> For more on the link between character and morality, see Frierson, 2006.

religious fanatic who, in the previous year, had claimed to have visions and became known around Königsberg as the “goat prophet.”<sup>7</sup> In that essay, Kant describes the “*fanatic* (*visionary, enthusiast*) [*Fanatiker* (*Visionär, Schwärmer*)],” who is

properly a deranged person with presumed immediate inspiration and a great familiarity with the powers of the heavens. Human nature knows no more dangerous illusion. If its outbreak is new, if the deceived human being has talents and the masses are prepared to diligently accept this heaven, then even the state occasionally suffers raptures.

Enthusiasm leads the exalted person to extremes, *Muhammad* to the prince’s throne and *John of Leyden* to the scaffold. (2: 267)

During the following two years, Kant writes his *Dreams of a Spirit-Seer*, a detailed investigation of the claims of the visionary theologian Emanuel Swedenborg. And throughout all of his discussions of mental disorder, Kant includes accounts of religious enthusiasm, or *Schwärmerei*. In *Anthropology*, for example, he worries about “oracular outbursts” (7:145), explicitly mentioning Socrates but no doubt also including those in Kant’s own day who claimed privileged access to divine insights. This sort of “enthusiasm [*Schwärmerei*]” where “certain judgments and insights are assumed to spring directly from inner sense (without the help of understanding), and . . . are further assumed to command themselves, so that sensations count as judgments . . . stands in close relation to derangement” (7:145). Later Kant is even more explicit, explaining that “*Schwärmerei* . . . is mental illness: the tendency to accept the play of ideas of inner sense as experiential cognition, although it is only a fiction” (7: 161). And elsewhere Kant compares “*Superstition* . . . with dementia, *Schwärmerei* with insania” (7:203).

For Kant, the association of *Schwärmerei* with mental illness has important implications for both epistemology and ethics. With respect to epistemology, Kant orients his *Critique of Pure Reason* as the only means of “sever[ing] the very root of . . . *Schwärmerei* and *superstition*” (B xxxiv). But the *Critique* accomplishes this by an analysis of the conditions of possibility of experience, understood to include both scientific empirical judgments and ordinary empirical judgments about the world. If individual prophets or seers are capable of supernatural experiences, of making empirical judgments about realities beyond ordinary human access, then Kant would need to provide an account of the conditions of possibility of such experiences. But of course, such an account would undermine the way in which Kant’s epistemology seeks to limit the scope of human knowledge. Thus Kant needs to place supernatural experiences in a category outside of the human cognition for which his philosophy provides conditions of possibility. By labeling *Schwärmerei* as a form of mental illness, Kant is freed from the need to provide a philosophical defense of it, and Kant’s critique can even provide a crucial *antidote* to it.<sup>8</sup>

Morally, the isolation of *Schwärmerei* as mental illness is arguably even more important. Kant even proposes “to recommend to the police that morality must not be allowed to get too close to” *Schwärmerei*. (11: 142). For Kant, the moral law must be autonomous; it is a law that human reason gives to itself. But religious enthusiasts often claim privileged access to specific moral or religious requirements of God.

The course of things is roughly this. First *genius* is very pleased with its bold flights, since it has cast off the thread by which reason used to steer it. Soon it enchants others with its triumphant pronouncements and great expectations and now seems to have set

---

<sup>7</sup> See the translator’s introduction to “Maladies” in Kant, 2008.

<sup>8</sup> See too 11: 142-3, where Kant specifically prescribes education about how to properly conduct science as a means for warding off *Schwärmerei*.



itself on a throne which was so badly graced by slow and ponderous reason . . . then its maxim is that reason's superior lawgiving is invalid – we common human beings call this *enthusiasm* [*Schwärmerei*], while those favored by benevolent nature call it *illumination*. Since reason alone can command validly for everyone, a confusion of language must soon arise among them; each one now follows his own inspiration, and so inner inspirations must ultimately be seen to arise from the testimony of preserved facts, traditions . . . – in a word, *superstition*. (8:145)

In *Religion*, *Schwärmerei* is more specifically tied to a supposed insight into the operation of grace in one's life, but, as throughout his discussions of *Schwärmerei*, Kant continues to see it as both something that keeps us “distant from the good based on . . . a well-ordered conduct of life” (6:83) and as a “kind of madness” (6:173).

#### d) Madness and egoism

Finally, an overarching purpose that unites all of Kant's discussions of mental disorder is its connection with “egoism.” Egoism is a general feature of civilized human beings. As Kant explains, “From the day that the human being begins to speak by means of ‘I,’ he brings his beloved self to light wherever he is permitted to” (7:128). Generally speaking, egoism is any absolute prioritizing of the self over others, but it takes three forms, at least two of which are important for understanding mental disorder. First, there is “the *logical egoist*, [who] considers it unnecessary to test his judgment also by the understanding of others” (7:128). Kant clearly means “understanding” here in the broad sense to include all higher powers of the cognitive faculty, and he even extends this further when explaining the opposite of egoism, where we do measure ourselves against others, and do so even in the case of our *senses*, “for example, whether a ringing is merely in our ears or whether it is the hearing of bells actually being rung” (7:129). A second form of egoism is “the *aesthetic egoist*,” who “is satisfied with his own taste” regardless of the opinions of others (7:129-30). Finally, “the *moral egoist* limits all ends to himself, sees no use in anything except that which is useful to himself, . . . and puts the supreme determining ground of his will simply . . . in his own happiness, not in the thought of duty” (7:130).

Although Kant insists that what all forms of madness have in common is egoism (7:219), the inverse is not true; in general, egoists are sane. One can hold one's own opinions to be correct without regard to the opinions of others and can prefer one's own interests while ignoring moral requirements to attend to others' without being mentally disordered in any way. Although Kant's moral philosophy emphasizes that virtue is the only life that accords with pure practical reason, he adamantly reserves the term “evil,” *rather than* “mad” for those who reject the demands of morality for the sake of their own happiness. And although the failure to measure one's judgments against those of others will likely lead to manifold errors, error is not, for Kant, equivalent to mental disorder; and Kant does not think that egoism always involves or even leads to mental disorder.

But egoism can contribute causally to mental disorder, and, more importantly, mental disorders draw attention to the nature and dangers of egoism in a particularly poignant way. Mental derangement is an extreme form of theoretical egoism, where one is *incapable* of comparing one's reason or judgment or even sense perceptions with those of others. Hypochondria and melancholia, similarly, are mental disorders that lead one to regard one's own inner sense as the basis for judgments about the world, without measuring those judgments against those of others. And passions are extreme forms of moral egoism, where one goes beyond sacrificing morality to self-interest and sacrifices both morality *and* self-interest to some one particular passion with which one is obsessed. (Hence in this case,

one can be *both* properly evil *and* subject to mental disorder, since the illness is itself an extreme form of moral egoism, which is evil.) Thus mental disorders reflect egoism taken to an extreme. Or, conversely, egoism is a sort of mental disorder under control, a mental disorder that leaves its patient functional, even thriving, but divorced from the community that is necessary for truth, beauty, and goodness.

### III. Implications for Kant's philosophy

Kant is best known for his *Critique of Pure Reason* (the “first *Critique*”), according to which the world that we experience is a world in part constituted by our own mental structures, and for his moral philosophy, according to which human beings are free agents who ought to act in accordance with a “categorical imperative” that prescribes choosing in accordance with principles of action that can be made universal. The fact that people can be afflicted with mental disorder has the potential to undermine both aspects of Kant's project. In this section, I briefly highlight the implications of Kant's own account of mental disorder for these key aspects of his philosophical project.

As I noted in the last section, one of the most important reasons for Kant's account of mental disorder is its role in combating religious *Schwärmerei*. To some extent, this aspect of Kant's account fits well with both his theoretical and practical philosophy. The first *Critique* aims to “sever the very root of . . . *Schwärmerei*” (Bxxxiv), and one of the main forms of “heteronomy” that Kant opposes in his moral works is any form of morality that ties the authority of moral rules to their foundation in “the will of God” (5:40, see too 4:443).

But Kant's most extreme account of mental illness – the “positive unreason” that constitutes craziness (7:216, cf. 2:268) – also raises deep philosophical problems for precisely the philosophy that is meant to combat religious extremism. As I noted in my introduction, Bertrand Russell, in an important critique of Kant, wrongly claimed that the possibility of radically different human cognitive structures never occurred to Kant. Kant's account of mental disorder makes clear that human beings are susceptible to a wide variety of different mental structures. Russell was particularly interested in the effect of these differences on necessary truths of mathematics, in which context Kant's insistence that one can be afflicted with a “positive unreason” in which one imagines “the squaring of the circle” (7:216) is particularly telling. In the context of Kant's concern with religious *Schwärmerei*, it is perhaps more important that one's craziness can lead one to “the unveiling of the supersensible forces of nature, and the comprehension of the mystery of the Trinity” (7:216). And for one with an “inability to bring one's representations into even the coherence necessary for the possibility of experience” (7:214), the efforts of the *Critique of Pure Reason* to elucidate “conditions of the possibility of experience” are at best a pointless venture.

These problems are made particularly acute by the fact that Kant does not seem (at first) to have a basis for arguing that the mentally ill are *wrong*. In his *Anthropology*, Kant insists that these conditions are diseases, and even if he does not think that they can be prevented or cured, Kant certainly does not consider them healthy forms of human life. Kant insists that such patients are afflicted by “delusion” (7:215), “derangement,” and “madness.” But Russell seems correct that Kant's *Critique of Pure Reason* undermines any legitimate basis for this critique. If one who is mentally ill literally lives with “a totally different standpoint” (7:216), and if, as the first *Critique* insists, “objects must conform to our cognition” (Bxvi), then Kant's claim that such a person lives “in his own world” (7:119) must be understood *literally*. But then the mentally ill are not really *ill*; they simply experience a different world.

In a sense, Kant is willing to accept this conclusion. He does describe the mentally ill as living in a different world, and he allows that “the powers of the unhinged mind still arrange themselves in a system” and even arranges cognitions according to “a principle of unity” (7:216). As he writes in a marginal note in his *Anthropology*, “There is a system in lunacy.” Nonetheless, Kant insists that in the insane, “the faculty of thought . . . is not working *objectively* toward *true cognition of things*” (7:216). And Kant’s attitude throughout his discussion of mental disorder is very far from the idealized late medieval conception of “madness [as] elements of a difficult, hermetic, esoteric learning” that constitutes “knowledge.”<sup>9</sup>

Kant never explicitly answers Russell’s objection. He never explicitly discusses the way in which his account of mental disorder is compatible with his first *Critique*, nor even what the implications of each would be for the other. But Kant’s texts rule out certain possibilities and, I think, strongly suggest a Kantian resolution to Russell’s challenge. The first *Critique* rules out a sort of objectivist realism that bases truth claims about the world on knowledge of what the world is like independent of our structures of cognition. And Kant’s account of mental disorder – combining an admission of diversity in cognitive structure with a strong normative insistence on the divergence between madness and truth – rules out a subjectivism that makes claims about the world true if those claims are grounded in one’s individual perception of the world filtered through one’s own cognitive structures.

*Anthropology* also includes a crucial clue to resolving the apparent conflict between the possibility of madness and the strong claims of the first *Critique*. Madness is an extreme form of egoism, and egoism, it turns out, is antithetical to truth. Kant suggests this in his anthropology (7:128-9), but makes it explicit in his lectures on logic:

[Logical] egoism is . . . when one holds that he alone judges rationally, that no one else is in a position to judge something or better to be able to have insight into it. . . . [But] one cannot be certain whether one has judged rightly or not if one has not compared his judgments with the judgments of others and tested them on the understanding of others. For a cognition is not correct when it agrees with my private understanding but when it agrees with the universal laws of the understanding of all men. (Bloomberg Logic 24:187)

The “touchstone of the understanding” is precisely the comparison of one’s own judgment with “other men’s insights” (Bloomberg 24:178). Hence in “What is Enlightenment?”, even in the midst of defending the importance of using “one’s *own* understanding” (8:35), Kant insists that only a “*public* use of one’s reason . . . *as a scholar* before the entire public of the *world of readers*” can bring about enlightenment (8:35, 37, emphasis original).

What Kant’s account of cognitive disorders shows, fundamentally, is the *social nature of truth*.<sup>10</sup> The “experience” for which Kant’s first *Critique* provides the conditions of possibility is not the experience of isolated individuals. Kant consistently uses the first person *plural* in the first *Critique*: “objects must conform to *our* cognition” (B xvi) or “objects conform to *our* way of representing” (Bxx) or “all *our* cognition begins with experience” (B1). These are not, it turns out, mere stylistic features of his work. Nor does Kant intend the work to apply only to a small “club” of philosophers covered by his magisterial “we.” Rather, the point of the use of the first person plural in the first *Critique* is tied to Kant’s

---

<sup>9</sup> Foucault, 1965, 21

<sup>10</sup> In this respect, Kant’s philosophy is crucially distinguished from that of Descartes’s *Meditations*, a philosophy that begins with an isolated “I” that knows only itself, and from the first part of Hume’s *Treatise of Human Nature*, which is similarly solipsistic.

long-standing concern with egoism. The only sort of *Critique* that can truly ground objective knowledge of the world must be a *social Critique*, one that can itself be criticized and refined by others. Kant's philosophy, in order to be properly objective, depends upon being comprehensible to his readers. And any account of the world, however true to one's own "experience," cannot truly be *experience* – that is, cannot be *objective* knowledge of a world – unless it can be shared. Cognitive disorders are disorders precisely because, by cutting one off from others, they cut one off from any truly objective world.

Cognitive disorders are not, of course, the only sorts of disorders that Kant highlights, and they are not the only disorders with profound implications for the rest of Kant's philosophy. Affects and passions, because they affect the faculties of feeling and desire, do not raise significant problems for Kant's theoretical philosophy, but they do problematize the transcendental freedom that underlies Kant's moral theory. In particular, Kant's moral theory depends fundamentally upon human freedom, but his account of affects and passions seems to undermine this freedom.

Of the two mental illnesses, affects at first seem the most troubling for Kant's moral theory, since they bypass human choices altogether. As noted above, one under the influence of affect acts immediately; feeling simply leads directly to action without any intermediary deliberation. Because they are feelings that preclude even the most minimal reflection upon one's actions, affects even seem to contradict Kant's claim that humans' "power of choice . . . cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim*" (6:23-4). Henry Allison has concisely summarized the general point here – "incentives do not motivate by themselves causing action but rather by being taken as reasons" (Allison, 1990, 51) – and Allison and others<sup>11</sup> have put this "Incorporation Thesis" to great effect in interpreting Kant's moral theory. Given how central the incorporation thesis is for contemporary accounts of Kant's ethics, affects – where a person is motivated by a sensuous incentive without any deliberate "incorporation" into a maxim – seem to undermine Kant's ethics at its root.

In fact, however, Kant addresses affects simply by allowing for their presence. For most human actions, and for every action for which a person can be held responsible, the incorporation thesis holds. But Kant's famous statement of the incorporation thesis in *Religion* (above, 6:23-24) is not a general claim about everything that human beings do. Rather, Kant is making an observation about "freedom of the power of choice" (6:23). And while most human actions are free acts of the power of choice, not all are. When one acts out of affect, one's deliberative faculty – one's power of choice – is not involved, and thus one need not incorporate one's incentives into any maxim. It is precisely for this reason that affects "can indeed coexist with the best will" (6:408). And by allowing this possibility for human beings, Kant may diverge substantially from recent interpretations of him, an account of human behavior that allows for non-deliberate action does no real damage to his moral theory – since most behavior is still deliberate – and is, arguably, a better fit with actual human life. Allowing for the possibility of affects in this way accommodates mental illness in a way that restricts the scope of Kant's moral theory without moderating its demands in human choices.

The bigger potential problem for Kant's moral theory, then, is the presence of passions. Passions allow for instrumental reflection on how to pursue their objects, but they

---

<sup>11</sup> The use of the incorporation thesis by Henry Allison (1990) has found its way into virtually all recent interpretations of Kant. See, for example, Baron, 1995; Korsgaard, 1997; and Reath, 2006.

do not allow for the sort of higher order reflection needed to compare the object of one's passion with the moral law or with other possible objects of desire. Here one might think that a person afflicted by a passion cannot be held responsible for their actions, since they do not compare their ends with other possible ends, but the passions are nonetheless located in one's choices, so Kant cannot simply exclude these cases from moral evaluation as he does with affects. And in fact, Kant does not exclude the passions, but terms them "properly evil" (6:408). Kant is able to do this because although the passions do not specifically involve comparison of their objects with the moral law, they nonetheless involve wholehearted commitment to the pursuit of their objects *without* consideration of a moral law *that one could consider*. However dismissive of every other concern one with a passion is, his conscience "follows him like a shadow . . . . He can at most . . . bring himself to *heed* [the moral law] no longer, but he still cannot help *hearing* it" (6:438).

Like insanity and dementia, the passions represent a sort of mental illness that has its roots in egoism. Kant discusses a "moral egoism" that "limits all ends to himself, sees no use in anything except that which is useful to himself" (7: 130). Unlike *this* sort of moral egoist, one under the influence of passions does not focus on what is useful to himself overall. In fact, Kant's account of passions often emphasizes that they undermine not only morality but even one's own overall happiness. But one under the influence of passions is like the egoist in putting his particular inclinations ahead of any comparison with a law that could be universal. In fact, one under the influence of passions might even be seen as a sort of ultra-egoist, focusing on a particular inclination not only to the exclusion of its compatibility with the interests of others but even without considering his *own* other inclinations.

Unlike logical egoism, the practical egoism of the passions is not the result of isolating oneself from community with others. In fact, Kant insists that passions actually "assail his nature . . . as soon as he is among human beings" (6:94). Human beings in the state of nature have simple needs that are easily satisfied, and because they are not motivated by comparison with others, passions have no opportunity to arise. But in community with others, the very interaction and comparison with others that helps prevent cognitive disorders can actually become a cause of passions. In his earliest writing on mental illness, in fact, Kant seems even more pessimistic about human society. In "Maladies," he claims,

The human being in the state of nature can only be subject to a few follies and hardly any foolishness . . . . From where should he draw the material for foolishness, since, unconcerned about another's judgment as he is, he can be neither vain nor inflated? . . . . Which dementia can well befall him since he never has cause to venture far in his judgment? Insanity . . . is surely wholly and entirely beyond his capacity. . . . The means of leavening for all of these corruptions can properly be found in the civil constitution, which, even if it does not produce them, nevertheless serves to entertain and aggravate them. (2:269)

On the one hand, as we have seen above, it seems as though society can alleviate cognitive disorders. On the other hand, as Kant emphasizes here, society seems responsible at least for passions and arguably for the possibility of any mental disorder at all.

For Kant, then, the egoism that can lead to mental disorder is not the egoism of the state of nature. The isolated person in the state of nature, like Rousseau's noble savage, does not have egoism in the proper sense. Lacking real self-consciousness, the person in the state of nature is happily occupied with immediate experiences and simple desires. The egoism that characterizes mental derangement is an egoism that arises only in the context of society, an egoism that comes when one has a sense of self developed through comparison with

others, but turns wholly inward, trusting one's own authority over "experiences" that are far from the state of nature in which the "savage" finds himself.<sup>12</sup> Theoretically, one might avoid all mental disorder through returning to the state of nature. Like Rousseau, however, Kant holds that such a return is neither possible nor desirable (see 7: 326). For one with a developed self-consciousness, any attempted return to the state of nature would only exacerbate mental derangement. The real means of preventing mental disorders is to oppose egoism with "*pluralism*, that is, of thinking in which one is not concerned with oneself as the whole world, but rather regards and conducts oneself as a mere citizen of the world" (7:130). And, as Kant immediately adds, "This much belongs to anthropology" (7:130).

In other words, what seem to be philosophical problems arising for Kant from the problem of mental disorder become further reasons for Kant's pragmatic anthropology. By teaching his students and readers to be engaged citizens of the world, Kant aims to cut off the most severe forms of mental illness. And by drawing attention to the nature of mental illness, Kant can provoke his students to become such citizens. At the same time, this pedagogical purpose of Kant's account shows how *important* the project is for Kant's Critical philosophy. Kant's pragmatic anthropology as a pedagogical task is necessary in order to create the social world within which the "we" of his first *Critique* finds an audience, and within which the world-oriented self-mastery of his students and readers can find a home in the universality of the moral law. Kant's theory of mental disorder, at first apparently a tangent to or even a problem for his Critical philosophy, turns out to be an essential part of actualizing that project in a social world.

#### IV. Conclusion: Some Conjectural Implications for Psychiatry<sup>13</sup>

Kant's theory of mental disorder was developed in the context of an anthropology that Kant hoped would displace that of Ernst Platner. Kant's hopes were unfulfilled. As John Zammito has pointed out,

Platner's book became the text for a number of anthropology courses that developed over the last quarter of the [18<sup>th</sup>] century in Germany . . . . Platner . . . proved extraordinarily important for the emergence of empirical anthropology in the late 18<sup>th</sup> century in Germany and – what needs underscoring – for a generation it was they and *not* Kant to whom the discipline looked for orientation. (Zammito, p. 253)

In the 19<sup>th</sup> and especially 20<sup>th</sup> century, Kant's emphasis on a pragmatic and non-physiological approach to the study of mental disorder has, if anything, been even more sidelined.

To some extent, Kant would be pleasantly surprised by the development of neurobiological accounts of mental disorder. Although he refers to "inquiries as to the manner in which bodily organs are connected with thought" as "eternally futile" (10:146), Kant is willing and even eager to appeal to physiological treatments when they are reliable and available (see 7:213, 220). When Kant insists that philosophers rather than medical doctors should be determining sanity in juridical contexts, it is because of the undeveloped state of medical science (7:214). Even when Kant objects to physiological approaches to anthropology, he refers to *both* approaches that emphasize bodily bases of mental states and,

---

<sup>12</sup> An important implication of this view in Kant is that even his most biological accounts of the origin of mental illnesses as "hereditary" (7: 217) must be tempered by the fact that certain social preconditions are necessary to activate these "germs of madness" (7: 217).

<sup>13</sup> My suggestions here regarding contemporary psychiatry are speculative at best. Not being an expert in contemporary psychiatry, my hope in this article is more to prompt further reflection among those more familiar with contemporary practice than to defend in detail any particular claims in that respect.

more importantly, to approaches that cannot be put to any practical use. For Kant, the former approach *implies* the latter because of the limited knowledge of how to manipulate the body to bring about shifts in mental states. Thus Kant says,

He who ponders natural phenomena, for example, what the causes of the faculty of memory may rest on, can speculate back and forth . . . over the traces of impressions remaining in the brain, but in doing so he must admit that in this play of his representations he is a mere observer and must let nature run its course, *for he does not know the cranial nerves and fibers, nor does he understand how to put them to use for his purposes. Therefore* all theoretical speculation about this is a pure waste of time. (7:119, emphasis added)

Especially with respect to those mental disorders that are particularly extreme or untreatable by non-physical means, any progress in medical-physiological science that could prevent, cure, or alleviate the ill effects of such disorders would certainly be welcomed by Kant (15:943, 947).

But to another important extent, Kant's approach to mental disorder, rooted in a *Weltweisheit* at once philosophical and pragmatic, provides an important counterweight to the clinical approaches that had already begun in the 18<sup>th</sup> century and have developed even further today. For Kant, "Medical science is *philosophical* when the sheer power of man's reason to master his sensuous feelings by a self-imposed principle determines his manner of living" (7:101). Kant's concern with physiological approaches to mental disorder is not merely that they do not work. Such approaches also put one's mental life in the hands of someone else. Rather than taking charge of one's own mental well-being, one "has a doctor who decides upon a regimen for me" (8:35). And this turning over of one's own mental capacities to another grates against the autonomy that Kant repeatedly emphasizes in both morals (see especially *Groundwork* and *Critique of Practical Reason*) and intellectual life (see "What is Enlightenment?").

Moreover, Kant's *Anthropology* contains hints that he would be extremely concerned about the implications for freedom of the institutionalization of mental disorder. He repeatedly points out the ways in which even those with mental disorders can be relatively functional members of society, such as how one without judgment can nonetheless be productive in matters of understanding (7:204). And he insists, with respect to the serious mental illness of dementia,

I have never seen anyone who has been cured of this disease (for to rave with reason is a special predisposition). However, they are not to be reckoned among the hospital buffoons; for, being concerned only with themselves, they direct their supposed craftiness only to their own preservation, without putting others in danger, and therefore do not need to be locked up for reasons of safety. (7:215)

Even while insisting that mental disorders can be quite serious and even incurable, Kant still does not think that society needs to impose its own order on those suffering from them. In many cases, Kant would prefer to leave those with mental disorders their autonomy, as long as they are not "putting others in danger."

Kant saw his anthropological discussion of mental disorder as a philosopher's supplement to medical treatments of mental disorders. Those medical treatments, from Kant's perspective, suffered from the deficient state of medicine at the time. And in that respect, Kant's account of mental disorder is outdated. But they also suffered from more intrinsic problems. Ultimately, for Kant, mental faculties (and thus mental disorders) are the province of psychology rather than physiology; physiologists can at best find the physical underpinnings of phenomena described and differentiated psychologically. Thus any

physiological account of mental disorder must depend upon something like the psychological account that Kant proposes in his *Anthropology*. But Kant's *pragmatic* anthropology also has particular advantages over even a well-developed physiological account of mental disorder. Specifically, by shifting attention to ways that one's life-style and self-discipline can work to prevent, ameliorate, and cure mental disorder, Kant provides individuals with a way of autonomously working for their own mental health. By showing dangerous parallels between mental disorder and moral failings such as egoism and religious fanaticism, Kant can show how failures of autonomy in ordinary life both resemble and lead to mental disorder. And finally, by being open to the fact of serious cognitive illness, Kant can make the profoundly important philosophical point that the world that we live in is fundamentally a social world that depends upon our *cooperative* autonomy.