Lubomira Radoilska. Autonomy in psychiatric ethics. Penultimate version December 2014. Final version published in: Sadler, J., van Staden, G. W. and Fulford, K. W. M. eds. *The Oxford Handbook of Psychiatric Ethics*. Oxford: Oxford University Press, 2015, pp. 354-371

Should autonomy play a significant role in psychiatric ethics? Is it even a legitimate normative concept? This chapter will explore four kinds of skepticism about autonomy in general and its applicability to psychiatric ethics in particular. It will be argued that although there are valuable lessons to be learnt from each of these skeptical challenges, their overall contribution is best understood in terms of friendly correctives to an autonomy-centered normative and conceptual framework instead of viable alternatives to it. The argument will proceed as follows. Sections 1-4 will provide each a logical reconstruction of a distinct skeptical line of reasoning about autonomy and expand on its implications for psychiatric ethics: (1) skepticism about personal autonomy; (2) skepticism about autonomy as an agency concept; (3) vulnerability-grounded skepticism about autonomy; and (4) paternalism-friendly skepticism about autonomy. The penultimate section 5 will identify and explore the underlying presuppositions that motivate the previously discussed forms of skepticism about autonomy. Finally, the concluding section 6 will reflect on the significance of psychiatric ethics for rebutting skepticism about autonomy and developing a new, more promising positive theory.

1) Skepticism about personal autonomy

In her inaugural address to the Aristotelian Society (2002), titled 'The Emperor's New Clothes', Onora O'Neill presents a sustained argument against what she identifies as misguided overreliance on autonomy in recent normative discussions. According to O'Neill, this unwelcome development can be explained, as it were, in terms of conceptual carelessness: work on autonomy has typically, though not universally failed to distinguish between three conceptions of autonomy – individual, rational and Kantian – whose nature,

scope and normative significance could not have been more different. While the latter, Kantian conception of autonomy applies to principles of action, the former two conceptions, individual and rational, are predicated either to individual selves or to individual choices, viz. psychological processes that have led to these choices. More importantly, these conceptions express different, and often conflicting, clusters of normative considerations. For instance, the autonomy of the will, which Kant presents as 'the sole principle of all moral laws and of duties in keeping with them' (Kant 1787: 5:33), can only be such a principle to the extent that it takes the form of a universal law, a maxim of action whose validity owes nothing of substance to the fact that it happens to be concerned with an action of mine rather than yours, or any other particular individual's. In other words, Kantian autonomy's centrality for ethical reflection comes hand in hand with a deliberate rejection of individuality as a morally significant consideration. On this view, projects, whose import stems from being a person's own in the sense of embodying her unique individuality, her authentic take on the world, would be morally suspect at best. Such projects would hardly be called for to illustrate instances of autonomy as self-determination. If relevant at all, they could only provide negative exempla – of self-indulgence, the lawlessness, or heteronomy of will.

This concise reconstruction suffices to make the contrast between Kantian and individual autonomy particularly salient. For, being a person's own in the substantive way precluded by the Kantian conception is a recurrent defining feature of autonomous actions and choices on the individual conception (e.g. Dworkin 1988; Frankfurt 1998). This is because the cluster of normative considerations that this conception is meant to capture has strong political underpinnings, aptly expressed by the title of an influential collection of essays on autonomy from the late 80s: the Inner Citadel (Christman 1989). Individual autonomy is thought of in opposition to coercion and undue influence, on the one hand, and on the other, internal obstacles to meaningful self-determination, such as ambivalence and self-deception. Drawing

on the political metaphor of the self as inner citadel, it becomes apparent that the individual equivalent of Kantian heteronomy should be twofold, corresponding to the origins of possible threats to autonomy, which can be either external or internal. In the first instance, individual heteronomy results from unfreedom to make up one's mind, i.e. the inner citadel is under occupation. In the second instance, individual heteronomy results from inability to make up one's mind, i.e. the inner citadel is brought to a standstill.

This inherent duality of individual autonomy, respectively heteronomy is not explicitly addressed in O'Neill's analysis. Instead, the primary target of criticism is the idea of 'mere independence' from outside that O'Neill takes to be at the heart of the approach under consideration (ibid., pp. 2, 4). The notion of inner obstacle that could belie the autonomy of related choices and actions is criticized in the context of the third and final conception of autonomy identified by O'Neill, rational autonomy. More specifically, the worry is that a secondary form of practical rationality, instrumental rationality, has been inappropriately placed at the center of autonomy discussions, dislodging in the process what should be central normative considerations. This is because instrumental reasoning is only subject to normative constraints that derive from the requirements for minimal consistency among a person's various ends and efficiency of the means that she chooses in the pursuit of these ends. In Kant's terminology, the form of this kind of reasoning is given by the so-called hypothetical imperative: as long as I am committed to a particular goal, I should be also committed to taking up the necessary steps to achieving that goal, or else I would be practically irrational. Clearly, this kind of rationality assessment does not extend to the ultimate ends of action: if I choose to abandon a particular goal of mine, I will no longer be required, on pain of irrationality, to undertake the means necessary to its realization. By insulating the ends of action from rational criticism in this way, the underlying instrumental conception of rationality demonstrates its neutrality, if not indifference to morality. For its

function, in a nutshell, is to ascertain whether a person is good at getting what she sets her mind to. Aptitude in this respect has no direct bearing on the moral status of one's actions and choices; what is more, such aptitude might be morally perilous in the absence of further normative constraints that, according to O'Neill, cannot but have the form of a categorical imperative.

We are now in a position to appreciate the full skeptical potential of the paper's thesis:

> "Recent conceptions of autonomy have no claim to be 'the sole principle of all moral laws and of duties in keeping with them', and their claims to Kantian ancestry are greatly exaggerated. We have been admiring a naked Emperor of questionable legitimacy." (O'Neill 2003, p. 1)

Individual and rational autonomy, the two poles of personal autonomy (Buss 2008; Taylor 2005), are criticised as normative imposters, taking over under the false pretenses of being related to the only legitimate conception, Kantian autonomy. That they would not have been able to command respect on their own right, without claiming a Kantian lineage, flows from the assumption that neither voluntariness,[1] nor practical rationality conceived in any other than Kantian terms could be normatively central.

This line of reasoning qualifies as skepticism about autonomy rather than a critique from within, commending one conception of autonomy as superior to others. The skeptical underpinnings of O'Neill's project become salient when we turn to medical contexts where autonomy is predominantly, if not exclusively conceptualized as belonging to persons and their choices rather than possible principles of action. Yet, as argued in an earlier paper (O'Neill 1984) this kind of autonomy is particularly unsuited to guide medical practice, for it

[1] On the links between voluntariness and independence from others, see in particular Feinberg (1986, pp. 254-261).

requires intact cognitive and volitional capacities precluded by the very notion, and reality, of being a patient:

> "Since illness often damages autonomy, concern to respect it does not seem a promising fundamental principle for medical ethics. Medical concern would be strangely inadequate if it did not extend to those with incomplete autonomy. Concern for patients' well-being is generally thought a more plausible fundamental principle for medical ethics." (p. 173).

The thought is that personal autonomy whether we conceive it in terms of individual independence viz. voluntariness or in terms of individual ability to make rational choices in light of one's values, projects and commitments offers not just an inadequate, but often misleading, if not downright counterproductive, framework for ethical reflection in medical contexts. Developed in greater detail in later works, e.g. O'Neill (2002), Manson and O'Neill (2007), the gist of the argument can be broken down into the following constitutive steps. Firstly, personal autonomy, in either rational or individual variance, cannot provide an independent source of normativity, outside the Kantian conception. Secondly, both individual and rational autonomy misrepresent the source of normativity they ultimately draw upon as having to do with personal uniqueness instead of universality of principles. Thirdly, the limited normative import of personal autonomy, concerns about mere independence of choice and means-ends rationality, is poorly served by respect for this kind of autonomy, for it amounts to respect for a person's well-functioning cognitive and volitional capacities. As a result, respect for personal autonomy might naturally lead to disrespect of persons, whose autonomy resources are depleted, e.g. by illness. This is because their choices would be constantly open to question as to independence and minimal rationality. Fourthly and finally, articulated in personal as opposed to principled terms, autonomy would be particularly harmful a principle to adopt in medical settings. O'Neill does not explicitly consider the

implications of her analysis for psychiatric ethics, it is compelling to conclude that the issues raised about cognitive and volitional impairments as obstacles to patients' personal autonomy would be even more to the point with respect to psychiatric than general medical practice. Arguably, the experience of some cognitive or volitional impairment is a constitutive feature of mental disorder as a problem that should be addressed within a medical context.[2]

Table 1: Varieties of Autonomy: Kantian or Principled vs. Personal, including Individual and Rational

Kantian autonomy: self-determination is conceived in terms of universal law. It applies to the principles or maxims of actions or the 'will'. It is opposed to heteronomy, a state in which the will is determined by individualistic, arbitrary desires. Individual autonomy: self-determination in terms of self-control. It applies to individuals and their choices. It is opposed to, on the one hand, control or coercion from others and, on the other, loss of self-control for a variety of causes, including physical and mental illness. Rational autonomy: self-determination in terms of rational choice conceived as consistency and efficacy in the pursuit of projects that a person choses for herself, be they worthwhile or not. It applies to the means rather than the ends of action as it derives from an instrumental conception of rationality. It is opposed to attachments to impossible or self-contradictory projects.

Table 2: Challenges to Personal Autonomy with Clinical Illustrations

Challenge 1: The focus on personal uniqueness as opposed to universal principles offers a poor guide to psychiatric practice.

[2] See Bolton and Banner (2012) for a compelling analysis of mental disorder as unmanageable distress.

Illustration 1: Treatment of people with personality disorders or in advanced stages of Alzheimer's would rely on separating and protecting an 'authentic' self from the 'unauthentic' or 'imposter' selves 'caused' by the illness.

Challenge 2: Respect for personal autonomy might lead to disrespect for persons whose capacity for personal autonomy is impaired.

Illustration 2: Refusal of life-saving treatment by people with depression would be considered exclusively on the basis of whether they can satisfy the relevant capacity test on this occasion.

2) Skepticism about agent autonomy

In a recent paper on "Mental disorder and the value(s) of autonomy" (2012), Jane Heal identifies and critically examines a form of thought which is implicit in discussions about what we, as a society, owe to people with mental disorder. This form of thought builds upon intuitions which link respect for a person with respect for this person's autonomy. In light of these intuitions, the issue of how to treat a person with mental disorder may seem to revolve around the question whether or not this person has the capacity for autonomy. However, Heal argues, inquiries that share this logical form are methodologically inappropriate and potentially unhelpful in answering either of the questions they put together: what we owe to people with mental disorder and what is involved in autonomy as a capacity. The reason for this is twofold. Firstly, the apparent consensus about autonomy as a capacity for self-determination that ought to be protected from interference by a corresponding right to self-determination is too shallow to ground a coherent course of action in terms of respect for autonomy. Even if we work with the assumption that autonomy is part of the Enlightenment

project, we face an important dilemma since we have to choose between a Kantian or rationality oriented and a Millian or well-being oriented take on the nature and significance of autonomy. Secondly, even if were to reach a substantive consensus on the concept of autonomy, it would arguably require an intricate array of mental capacities, outside the reach of at least some people with mental disorder. Getting clearer on what autonomy is will not help us find out what it means to treat these people respectfully. In this respect, as Heal suggests, we might be better off 'cutting out the middleman' (*ibid.*, p.4) and articulating legitimate concerns within psychiatric ethics without referring to autonomy.

At first blush, this line of critique might seem too closely related to skepticism about personal autonomy to constitute a separate strand in its own right rather than a new version of the same thing. This initial impression has to do with the fact that on both accounts autonomy is deemed incompatible with the volitional and cognitive impairments that are to be expected in the context of mental disorder. Yet, despite this point of convergence, these are two separate arguments levelled at what turns out to be different aspects, if not conceptions, of autonomy. For instance, Heal's misgivings about the role that autonomy could legitimately play in psychiatric ethics do not derive from a sweeping skepticism about the prospects of autonomy as a normative concept. This is because autonomy is conceived by Heal here as a putative ideal of agency that, though not well-articulated and rather confusing at present, could be 'leading to the emergence of new ways of acting, feeling, and thinking, ways which are appropriately structured and enabled by the use of the future concept [of autonomy]' (*ibid.*, p. 23). Unlike O'Neill's account of personal autonomy as morally insignificant, if not perilous, Heal's argument allows for a positive conception of agent autonomy, one that would ground forms of life, whereby currently irreconcilable though equally valuable commitments and projects could be jointly pursued. In other words, a carefully thought through concept of agent autonomy would enrich the spectrum of possible patterns of living since it would

contribute to resolving underlying tensions between apparently conflicting objectives, such as personal independence vs. intimate interdependence, risk-taking and excitement vs. safety and reliance on others. Following this line of thought, what a cogent concept of autonomy should be expected to bring about is the possibility of a sophisticated balancing act. In order to be taken advantage of, the hoped-for conceptual opening would require, as Heal puts it, 'the robust workings of complex psychological and social structures in which emotions, habits, thoughts, etc. intertwine in an intricate way' (*ibid.*).

This partial endorsement of autonomy as a potentially helpful normative concept in the future elucidates the initial challenge, according to which autonomy might not be as relevant to psychiatric ethics as it is often assumed. At the root of this form of skepticism is the contention that, as an ideal of agency, autonomy opens the door for – if it doesn't actively promote it – an implicit attitude, according to which persons are worthy of respect and protection only in so far as they are agents. From this perspective, acknowledging the potential of autonomy to sketch a possible ideal of agency for us, easing the strain between currently unbridgeable frames of reference – independence and interdependence – goes hand in hand with distrusting autonomy as guidance in contexts, where agency but not necessarily personhood is impeded, erratic, or absent.

Table 3: Agent Autonomy: Key features, Challenge, and Clinical Illustration

Key features: self-determination is conceived as an ideal of agency incompatible with volitional and cognitive impairments.
Challenge: Agent autonomy focuses on sophisticated, creative, and complex activities as opposed to activities perceived as basic, routine or simple. In psychiatric practice, such focus would be unhelpful.
Clinical illustration: Residential care for people with Alzheimer's or intellectual disabilities

might offer opportunities for artistic engagement, but fail to make room for their sexual expression.

3) Vulnerability-based skepticism about autonomy

In a recent essay, Martha Albertson Fineman defines vulnerability as 'a state of constant possibility of harm' (2008–2009, p. 11). On her view, vulnerability is universal to the extent that it reflects our 'embodied humanity' and 'the present potential for each of us to become dependent based upon our persistent susceptibility to misfortune and catastrophe' (*ibid*., pp. 9, 12). According to Fineman, the universal character of vulnerability makes it not only a viable, but also urgent alternative to autonomy as a core liberal value. At the heart of the proposal is the idea that autonomy cannot provide a helpful focus for our political institutions because, by trying to respect autonomy, we lose sight of the various kinds of possible harms to which we are open by virtue of being human. In contrast, the argument goes, if vulnerability replaces autonomy as a central concern, it would be possible to both improve overall protection against potential harms and boost individual 'resilience in the face of vulnerability' across society (*ibid*., pp. 2, 13).

It falls beyond the scope of the present inquiry to discuss the merits of the proposal which builds upon the above contrast between autonomy and vulnerability. Instead, it is the contrast itself that is of particular interest, for it opens up logical space for a distinct skcepticism about autonomy. This becomes apparent if we unpack the suggestion that vulnerability considerations are to replace or at least complement autonomy considerations if we are to successfully forestall a wide array of probable harms within society. Considering the direct link between respect for autonomy and the so-called Harm principle within some strands of the liberal tradition (Mill 1859; Hart 1963; Feinberg 1986), it is plausible to infer that there

are harms undetectable by this principle and vulnerability is a new conceptual resource able to redress the resulting oversight. Since the Harm principle states that one's freedom may be legitimately constrained only in order to prevent harm to others, we are left with two options concerning the scope of this new, vulnerability-based skepticism about autonomy. The first is to claim that there is need for protection beyond harm to others and this is where vulnerability comes into play. The second is to advocate a broader understanding of harms than it is implied by the Harm principle.

Before commenting on each of these options, it is worth getting clear about the conception of autonomy whose limitations vulnerability is supposed to remedy. In Fineman's essay, the rejected notion of autonomy is steadily associated with independence. In a similar vein, Joel Anderson and Axel Honneth critique the liberal tradition for implying a conception of autonomy as absence of any restrictions on individual freedom (2005, pp. 128–29). In doing so, they effectively equate autonomy with negative liberty as defined by Isaiah Berlin (1958). Like Fineman, Anderson and Honneth consider that the critiqued notion of autonomy follows from 'a misleading idealisation' of the human condition whereby individuals are depicted as independent, self-reliant, and self-sufficient (*ibid*.). Moreover, Anderson and Honneth concur in suggesting that a vulnerability informed view of interpersonal relationships could help resolve the difficulties brought by the liberal ideal of autonomy as independence.

The core argument for embracing vulnerability instead of autonomy that emerges from both Fineman's and Anderson and Honneth's discussions could be broken down into the following constitutive steps. Firstly, autonomy means self-determination. Secondly, self-determination excludes determination by others. Thirdly, following from the conjunction of the preceding two points, autonomy requires independence from others. Fourthly, human beings are not self-sufficient. In fact, it is implausible to imagine human life without various personal interactions which make mutual dependence both inescapable and valuable to us. Fifthly, a

relationship of mutual dependence is incompatible with the self-determination of the parties. Sixthly and finally, autonomy presents an impoverished and unattractive ideal of human life, while vulnerability is able to offer much-needed correction, for it acknowledges the significance of interdependence.

This overall conclusion brings vulnerability-based skepticism about autonomy closer to skepticism about personal than skepticism about agent autonomy. For it derives from a comprehensive concern about autonomy as a normative concept, concern that should arguably become even more pressing in the context of psychiatric ethics, where extreme patient vulnerability is often the case. This line of reasoning has been persuasively developed in a recent paper by Jennifer Radden (2012), who argues that consent might sometimes offer only inefficient, if not spurious, protection for either patients' best interests or right to self-determination in mental healthcare. More specifically, the discussion is focussed on the 'privacy stakes' for these patients, a term by which Radden refers to the likelihood that their confidentiality will be breached and the degree of harm that they would suffer from it. By reflecting on a series of relevant factors, including: the nature of mental disorder and therapeutic exchange, a normative framework which imposes conflicting professional obligations on mental health care professionals, and the persistent societal stigma associated with mental disorder, Radden concludes that there is a combination of high risk of disclosure and highly negative consequences, which place people treated for severe mental disorder "in a situation of extreme and continuing vulnerability" (*ibid.*, p.124).[3]

As significant as the links between skepticism about personal autonomy and vulnerability-based skepticism appear to be, the particular focus on the prevention of possible harms that remain undetected by the Millian Harm Principle – a distinctive feature of the latter, only

[3] It is important to note that neither Radden (2012) nor Anderson and Honneth (2005) draw radical sceptical conclusions about autonomy. Nevertheless, as shown by the preceding reconstruction, both papers may support such conclusions.

accidentally shared by the former – effectively brings vulnerability-based critiques of autonomy into a close contact with the fourth and final kind of skepticism about autonomy to identify and explore in this chapter:

Table 4: Vulnerability as a normative alternative to autonomy: Rationale, Related Challenge, and Clinical Illustration

Rationale: dissatisfaction with the Harm Principle as grounded in too narrow an understanding of harm and offering no protection beyond harm to others.
Challenge to Autonomy: autonomy presumes that independence from others is the default condition. In so doing, it provides insufficient protection to psychiatric patients, for whom dependence on trustworthy others is the default position.
Clinical illustration: Disclosure of confidential patient information in order to prevent possible harm to others, with unavoidably negative consequences for the patient. A patient with bipolar disorder at the manic stage of their disease might even insist on such disclosure and be in a position to satisfy the conditions of informed consent. On this occasion, their autonomy seems respected, but not their vulnerability.

4) Paternalism-friendly skepticism about autonomy

The shift from harmful interference by others to harm *tout court* that is implicit in vulnerability-based critiques of autonomy comes to the fore in recent justifications of paternalism. At the heart of these justifications lies the idea that there is an obligation to protect people from the harmful consequences of their own choices, if necessary, by curtailing these choices. This as we might call it paternalist intuition is well articulated in a recent paper by Danny Scoccia (2008) who argues that we should acknowledge two moral

principles, autonomy and beneficence. The former requires deference to other people's past and present values concerning how they should live their lives. The latter requires that we reckon what is good for others in a way that is temporally neutral with regard to present and future. More specifically, the autonomy principle tells us that it is wrong to use 'unsavoury means', such as coercion, manipulation and deception in order 'to thwart another person's self-regarding choices, except when: (a) he gave actual prior authorization for the coercion, or (b) he would now authorize it if he were calm, reflective, and well-informed'(p.358). In other words, we are not entitled to use a theoretical device, such as hypothetical rational consent. Respect for autonomy requires that we take a person's preferences and values as given, no matter how misguided or idiosyncratic they might seem to us. By contrast, beneficence is about the promotion of prudential values, such as health that would make another person's life go better from this person's evaluative perspective but conceived in temporally neutral terms.[4] More precisely, the beneficence principle 'identifies one's prudential good with the satisfaction of one's self-regarding desires or fulfilment of whatever views one happens to hold about what will make one's life go best. It implies that you can be mistaken about what's best for you over the long run because your preferences or views can change in ways that you fail to anticipate' (p. 362) According to Scoccia, autonomy and beneficence are equally fundamental as ethical principles, each representing a Rossian ''prima facie'' duty (p. 363). Hence, they are bound to clash on occasion. When they do, beneficence sometimes overrides autonomy.

To clarify the skeptical import of this view, two related features need to be considered. The first is the claim that the paternalist intuition as articulated above is implicit in common sense morality. The second is the claim that hard, as opposed to soft, paternalism reflects better ordinary moral intuitions and, moreover, fares better at the level of philosophical theory.

[4] On the nature of health as a value, see Radoilska (2009).

Following Feinberg (1986, pp.12–16), Scoccia understands soft paternalism as a conception, according to which the obligation to protect people from the harmful consequences of their choices only relates to cases, in which these choices do not appear to be fully or even sufficiently voluntary. Examples include instances of uninformed consent or refusal to consent, as well as decisions made in the absence of decisional capacity.[5] In this respect, soft paternalism is consistent with the view that autonomy is more fundamental a principle that beneficence: the duty to protect a person against the potentially harmful consequences of her own choices becomes morally appropriate only in cases where these choices cannot be considered as her own in a meaningful, sufficiently robust sense (Radoilska 2012). In other words, soft paternalism does not really advocate the priority of beneficence over autonomy; instead, it identifies cases where only beneficence, but not autonomy can be a relevant consideration.

In contrast, hard paternalism, the position Scoccia defends, violates – albeit permissibly – the principle of autonomy, for it interferes with fully voluntary and substantially self-regarding choices. Moreover, it does so by employing the so-called unsavory means of coercion, manipulation and deception. This radical departure from autonomy is deemed as appropriate in cases where the violation of self-determination alone and no other, more substantive right, such as freedom of religion will bring significant prudential benefit to the subject of paternalist interference (pp. 371–72).Thus, Scoccia excludes forced blood transfusion to adult Jehovah's Witnesses as unacceptable, for it would violate a more specific and substantive right than the right to self-determination. In this respect, hard paternalism is distinguished from moral paternalism which, by violating specific and substantive rights, such as freedom of religion, is unlikely to prudentially benefit the subjects of interference (p. 379).

[5] On the relationship between decisional capacity and alternative conceptions of autonomy, see Radoilska (2013b).

Hard paternalism is presented as particularly attractive in medical contexts. The central example of a prudentially justified paternalist intervention that has the additional benefit of matching ordinary moral intuitions is the denial of physician-assisted suicide to people who, like a suicidal sportsperson because of a minor injury that would prevent him from competing at the highest professional level, 'are better off alive than dead notwithstanding their belief to the contrary' (p.367). The thought is that, unlike moral paternalism, hard paternalism here does not ignore its intended beneficiary's evaluative perspective altogether. What it does is to counter the sportsperson's bias toward the preferences and values that he holds at present by adopting a temporary neutral perspective, from which preferences and values that he might plausibly endorse in the future weigh just as much as the actual ones.

We are now in a position to appreciate the skeptical import of Scoccia's paternalism. It comes as a natural consequence of a reductivist reconceptualization, according to which autonomy is not just one of two equally significant sources of moral obligation; more importantly, it is presented as lacking in substance of its own, an abstract right to self-determination whose sole interest is to provide a platform for specific substantive rights, such as freedom of religion.[6] The upshot is sweeping skepticism about autonomy as an agency concept. To see why this is so, suffice to follow the line of thought that leads, in the first instance, to conceiving of autonomy as reducible to the sum of specific interests that are, recognizably, best served, if left to individuals. For this initial step is what allows a proponent of hard paternalism, such as Scoccia to devise a theoretical framework where autonomy and beneficence are treated as opposing yet equally important normative principles. The risk of ending up with an unworkable set of conflicting demands is forestalled by providing a purely instrumental justification for either principle: autonomy matters and so thus beneficence; yet, neither matters in its own right. There are circumstances in which it is plausible to assume

[6] Another example Scoccia (2008) offers is the right to pursue intimate partnerships in tune with one's sexual orientation.

that people are best placed to act in their own best interests and this where autonomy takes precedence over beneficence. Alternatively, there are circumstances in which others might be more successful not just at promoting one's interests, but also at determining what these interests effectively are. The fact that Scoccia (2008, pp. 360–362) disavows an objectivist theory of value that would tell us what the good for everyone is, without the need of taking anyone's perspective on what makes a life worth living, does not help a great deal. Arguably, it makes things worse. We end up with an indefinite list of substantive interests that are worth protecting either under the principle of autonomy or under that of beneficence, whichever has better prospects of doing the job on a particular occasion. Yet, we are not given any reason for why the job needs doing in the first place. Like skepticism about personal autonomy, paternalism-friendly skepticism envisages that, ultimately, respect for autonomy amounts to respecting a kind of a wish list by isolating it from any form of critical engagement. While the former kind of skepticism objects to the idea of such a subjective wish list being at the heart of moral life, the latter kind of skepticism is happy to endorse it on condition that the wish list is organized in a particular way. Looking at the examples offered, it becomes apparent that what counts as a value or preference that fits the realm of autonomy rather than that of beneficence does not really flow from a person's evaluative perspective: a forced blood transfusion onto an adult Jehovah's Witness is deemed morally objectionable because it violates her right to freedom of religion; a physician-assisted suicide for a sportsperson who will not be able to compete at the highest level because of a minor injury is deemed morally objectionable because over time he might find new interests, other things worth living for. Why not anticipate that the Jehovah's Witness, just like the sportsperson, would develop new interests if she were to keep on living? Instead of addressing concerns about arbitrariness gaining a foothold in moral life that we saw motivating skepticism about personal autonomy,

paternalism-friendly skepticism amplifies these concerns by removing any principled way of (re)drawing the realm of autonomy.

Needless to say, paternalism-friendly skepticism about autonomy has an immediate and grave bearing on psychiatric ethics. If autonomy as right to self-determination can be easily overridden by beneficence in the presence of a secure capacity for self-determination, it becomes difficult to see how much protection this right could offer in the context of therapeutic encounters meant to address various kinds of agential vulnerabilities, whose feature in common may be described as a sense of insecure or inefficient capacity for self-determination, if not a breakdown of agency (Fulford and Radoilska 2012).

Table 5: Varieties of Paternalism: Hard vs. Soft Paternalism, Related Challenge and a Clinical Illustration

Soft paternalism: there is an obligation to protect people against the harmful consequences of their choices only if these choices are not fully voluntary and informed.
Hard paternalism: there is an obligation to protect people against the harmful consequences of their choices even if these choices are voluntary and informed.
Challenge to autonomy: there is no fundamental right to self-determination. Autonomy is just an umbrella term covering specific substantive rights, such as religious freedom. Unlike religious freedom, other specific rights might be overridden on grounds of beneficence.
Clinical illustration: The refusal of food by a person with anorexia and the refusal of a life-saving treatment by a person with depression would not seem to be on a par with the refusal of blood-transfusion by a Jehovah Witness in spite of having equally grave consequences.

5. Skeptical presuppositions

Having considered four major kinds of skepticism about autonomy and their implications for psychiatric ethics, we are able to identify three underlying presuppositions. The first is about the relationship between individual and joint agency. The second is about the nature of practical rationality. The third is about the significance of first-personal authority in deliberating about and planning for action. In the following, I shall critically explore each of these presuppositions in turn.

Presupposition 1: *Autonomy precludes joint agency*

An assumption shared by all four kinds of skepticism, albeit to a different degree, is that autonomy, as an ideal of agency, is at odds with a sound understanding of human interactions in the sense of doing things together as opposed to each individual agent doing things on his or her own, without impediment from others. Although intuitively appealing, this assumption is in fact mistaken. From the idea that individual autonomous agency requires a space where projects and commitments can be developed, pursued, revised, or abandoned 'at will', that is, without an agent having to ask permission from others, it does not follow that collective or joint autonomous agency is impossible. The correct entailment is the reverse: individual and joint autonomous agencies presuppose and facilitate each other. The opportunity to explore and engage freely in some projects of one's own goes hand in hand with various possibilities to co-opt fellow agents and to join in projects initiated by others on an equal footing. Nothing could be further from the circumstances of autonomous agency at either individual or collective level than the picture of disintegrated, atomistic social interactions where self-interest reigns supreme that some critics of personal autonomy are keen to depict as the price to pay for upholding such a misguided ideal (see in particular, Fineman (2008–2009), but also Oshana 2006 and Stoljar 2000). Looking at the preceding analysis, it becomes apparent that this picture fits the kind of allegedly beneficent intervention advocated by hard paternalism. For at the heart of this alternative conceptualization of agency lies the assumption that what

matters ultimately is the satisfaction of some significant individual interests, agents themselves being interpreted as more or less efficient means to achieving this ultimate end. In this respect, we could say, drawing on the famous analogy between personal ethics and state constitution in Plato's *Republic*, that politics is indeed moral psychology writ large. By precluding the so-called unsavory means of deception, manipulation, and coercion, autonomy also precludes a purely instrumental approach to fellow agents. In the same vein, the notion of privacy as a network of interceding spheres of action,[7] where individuals are sovereign, though not unlimited by others that animates the ideal of personal autonomy enables as much as it feeds from the reality of non-reductively joint autonomous agency.

Presupposition 2: *Personal autonomy implies an instrumental conception of practical rationality.*

A closely related assumption made by skeptics is to relate personal autonomy to an instrumental conception of practical rationality. This assumption is particularly salient in O'Neill's Kantian misgivings about personal, as opposed to principled, autonomy and Scoccia's defense of hard paternalism. Although arguing from very different and in many ways opposing starting points, both philosophers concur in concluding that its being inextricably associated with an instrumental conception of practical rationality makes personal autonomy, if not morally insignificant altogether, at least by far less central than it would seem at first sight. Though partly correct, this diagnosis is also somewhat misleading. Skeptics are correct to point out that within an instrumental framework of practical rationality personal, and for that matter moral, autonomy is bound to be peripheral. For such an instrumental framework posits as valuable certain desirable or actively desired states of affairs (Radoilska 2013a, pp. 120–124; Shapiro 2001). It is by virtue of being efficient in

[7] For a helpful discussion on how privacy might be more significant in psychiatric contexts than patient autonomy in terms of consent, see Radden (2012).

bringing about such states of affairs that both actions and agents are prized as practically rational. Importantly, these states of affairs are thought of as ends of action in terms of end results rather than instances of self-realization, or possible ways in which a person constitutes him or herself as an agent. That is to say, on the instrumental conception, agents cannot be ends in themselves on pain of practical irrationality; and so, no action or agent can be both practically rational and autonomous.

Nevertheless, skeptics are mistaken in assuming that personal autonomy is necessarily related to an instrumental conception of practical rationality. Taken together, hard paternalism and Kantian or principled autonomy do not exhaust the theoretical spectrum for defining the relationship between autonomy and practical rationality. If hard paternalism comes as a natural conclusion of endorsing a purely instrumental view of practical rationality, according to which all actions share one form, bringing about a desirable outcome, a formal view of practical rationality leading to a principled as opposed to personal conception of autonomy along Kantian lines is not the only alternative. A substantive view of practical rationality as intending and acting under the Guise of the Good (Radoilska 2013a, Tenenbaum 2007) is in a position to avoid both the heteronomy, if not anomy, that comes with an instrumental view and the impersonal character of a conception of autonomy that attaches primarily to universalizable principles of actions rather than individual agents. On this substantive view, autonomy is attributable to agents, such as persons, but also groups and living beings quite generally. In the concluding section of this chapter, more will be said about the advantages of endorsing such an agent-oriented conception of autonomy. For the moment, however, let us return to the third and final skeptical presupposition.

Presupposition 3: *An agent is not distinctly authoritative with respect to her actions.*

A powerful undercurrent that motivates all four kinds of skepticism about autonomy we considered earlier can be traced back to the idea that the perspective of an informed, yet

impartial observer is distinctly authoritative in ways in which an agent's first-personal perspective over her commitments, plans and actions can never be. The intuition at work here seems to be that since having a vested interest in the success of one's actions is constitutive of agency, agents lack sufficient critical distance toward their actions. In this respect, they find themselves in a position of epistemic disadvantage vis-à-vis the privileged kind of observer identified at the start of the discussion. This disadvantage can be construed as both evaluative and deliberative. The former expresses the idea that an informed but impartial observer would be better placed to assess the merits of a possible project than the agent, while the latter, that such an observer would also be better placed than the agent in figuring out what the best way is to ensure that the project, once undertaken, is a success. In both instances, the thought takes a similar form: the agent is much too involved to avoid bias; as a natural correction, a privileged observer's perspective must be recognized as significant and at least on a par with the agent's perspective. Yet, respect for autonomy as an agency concept demands that we take the agent's first-personal perspective as more authoritative than that of a privileged observer.

At the source of this skeptical line of thought we find the assumption that authority here should be thought of on the model of theoretical or expert authority. In other words, plans and actions are treated as facts to be known, interpreted, analyzed, and assessed. Their irreducibly practical dimension is ignored. Paraphrasing David Wiggins (1995) on truth in ethics, it is helpful to consider actions as practical truths whose nature is not only to be discovered but also to be created, if not ex nihilo, at least from scratch. Respect for autonomy as an agency concept requires us to acknowledge this kind of authority as belonging to agents as opposed to observers independently of whether the latter might be better informed, quicker reasoners etc. etc. That is to say, agential authority is practical, not theoretical (Tennenbaum 2009). By recognizing the distinctive nature of an agent's first-personal involvement in her actions, it

becomes possible to appreciate it as the foundations of a separate kind of authority that does not have to compete with, nor can be replaced by, that of an expert observer (Moran 2001). And so, instead of generating epistemic bias in need of further checks and balances, an agent's first-personal perspective opens up the possibility of practical, but also theoretical knowledge (Boyle 2009).

Table 6: Skeptical presuppositions and Possible Replies with Clinical Applications

Presupposition 1: Autonomy precludes joint agency
Reply 1: Individual autonomous agency presupposes non-reductive joint autonomous agency.
Clinical application 1: A meaningful therapeutic encounter enhances the autonomy of both patient and clinician.
Presupposition 2: Personal autonomy implies an instrumental conception of practical rationality.
Reply 2: The instrumental conception of practical rationality is inconsistent: it makes it impossible for an agent to be both autonomous and rational.
Clinical application 2: The efficient protection of a psychiatric patient's best interests is unadvisable whenever it relies on means that she finds alienating.
Presupposition 3: An agent is not distinctly authoritative with respect to her actions.
Reply 3: There is a distinction to be made between theoretical and practical authority. An agent might often lack theoretical authority over her actions, but still retain practical authority over them. The latter kind of authority is the more fundamental.
Clinical application 3: Loss of insight does not cancel out the practical authority a psychiatric patient has over her actions.

6. Rebutting skepticism about autonomy in psychiatric ethics

Having identified and critically explored the underlying presuppositions of general skepticism about autonomy, it becomes apparent why psychiatric ethics is an area where skeptics are particularly eager to make their case. According to the skeptic, the relative insignificance of autonomy as a normative concept should be immediately salient in instances of mental disorder where therapy is required. If autonomy is about letting people do their own thing, provided that they are competent and minimally efficient as agents, it cannot provide normative guidance for therapeutic encounters that imply the experience of vulnerability, pain, need, or distress. To put it crudely, the skeptical intuition seems to be that whenever assistance is appropriate, personal autonomy becomes obsolete. And so, patient autonomy is a contradiction in terms. On this view, the illusion of patient autonomy is maintained by the notion of service users, hinting at free choice not only of one particular treatment over another, but also free choice of whether to be treated at all.

Looking at some recent accounts of autonomy in psychiatric ethics, we should acknowledge that skeptics are onto something important here. For instance, many authors take autonomy to be at odds, if not downright incompatible with severe mental disorder. At the very least, instances of mental disorder are routinely presented as limit cases of autonomy.[8] A natural implication of this approach is to formulate issues in psychiatric ethics in terms of dilemmas, such as to empower or to protect (Dunn et al., 2008), to let people die or to respect their core values (McLean 2009). In this context, it is tempting to endorse what we may term local skepticism about autonomy as it applies only to psychiatric ethics as opposed to global skepticism about autonomy as a normative concept across a wide variety of domains. This is because local skepticism derives its appeal from the very dilemmas identified by some proponents of autonomy in psychiatric ethics: these dilemmas arise when the three skeptical

[8] For a critical discussion of this trend, see Fulford and Radoilska (2012) and Radoilska (2012).

presuppositions discussed earlier are tacitly endorsed and yet autonomy is expected to play a pivotal role in shaping normative expectations. From this perspective, rejecting the centrality of autonomy in psychiatric contexts in favor of another normative concept, such as beneficence might look like a good idea. Apart from resolving the aforementioned dilemmas, the benefit of adopting an explicitly skeptical view on autonomy in psychiatric ethics would be that, being localized, this kind of skepticism seems in tune with a broader recognition of autonomy as a central normative concept. For, as mentioned earlier, instances of mental disorder are typically treated as belonging at the margins of autonomous agency, if relevant at all (Jaworska 1999). Thus, debates on autonomy in psychiatric ethics often give the impression that the point of contention is whether an agreed general conception of autonomy should be extended to encompass this marginal domain of agency (Silver 2002; van Willigenburg and Delaere 2005). What is more, this shared conception of autonomy embodies the three underlying presuppositions that underpin skepticism about autonomy: (1) it focusses on individual as opposed to joint agency, (2) assumes an instrumental view of practical rationality, and (3) rejects the notion of agential first-personal authority. The upshot is global skepticism about autonomy as a normative concept. Once psychiatric ethics is conceived through the skeptical lens, the centrality of autonomy becomes questionable in every other domain of its application. This is because the dilemmas deemed central in psychiatric ethics are effectively due to the conjunction of skeptical presuppositions with a robust commitment to autonomy, which is incongruous. That is to say, skepticism about autonomy in psychiatric ethics is the precursor of a generalized attitude of skepticism, not a niche position. Hence, by acknowledging the central place of psychiatric ethics for theorizing about autonomy, we are able to revisit apparent dilemmas facing this vital area of normative ethics as well as address the very source of sceptical concerns about autonomy as a normative concept.

Following this line of thought, it becomes apparent that the dilemmas at issue are best resolved by rejecting all three skeptical presuppositions instead of downplaying the significance of autonomy in psychiatric ethics. For, as we saw in the preceding section, in spite of their joint appeal, neither presupposition withstands close critical examination. More specifically, these presuppositions, once clearly articulated, do not provide us with a compelling reason for relegating autonomy to the margins of normativity. What they do instead, is to rule out as unpersuasive a specific account that ties autonomy with individual as opposed to joint agency, unduly privileges instrumental at the expense of substantive practical rationality, and fails to recognize the distinctive authority of agents as more than sheer efficient causes (Radoilska 2013a, pp. 120–133). In this respect, a debt of gratitude is owed to skeptics. For such an account of autonomy, as tempting as it might appear to some at first blush, is confusing and should be rejected. Naturally, the rejection of an unhelpful account of autonomy doesn't warrant any skeptical conclusions about the prospects of this concept. On the contrary, it opens up space for more promising approaches, free from skeptical misconceptions. To correct these, a successful account should be able to flesh out an ideal of autonomy that informs constructive engagement with others as fellow agents, whose perspective on the world is authoritative in a practical sense. By moving away from a purely instrumental understanding of practical rationality, a promising conception of autonomy will effectively help revisit some of the most fundamental presuppositions about the nature of agency, transcending the reductivist view of agents as sheer efficient causes. Needless to say, psychiatric ethics will be the ultimate testing ground for new theories of autonomy. For, if agency in the context of mental disorder continues to appear as marginal or problematic in a way that leads to contrasting respect for autonomy with duties of beneficence, then skepticism has not been left behind.

7. Summary and Conclusions

This chapter identified and critically explored four kinds of skepticism about autonomy of relevance to psychiatric ethics. It was argued that the three presuppositions that underlie these skeptical approaches to autonomy, although plausible at first sight, do not withstand close examination. In conclusion, it was shown that patient autonomy is not a contradiction in terms but a vital normative concept that, once liberated from sceptical undertones, should play a pivotal role in both psychiatric and wider normative ethics.

Acknowledgements

I am grateful to John Sadler, editor of this *Oxford Handbook*, and an anonymous referee for their helpful comments and suggestions on an earlier version of the chapter.

References

Anderson, J. and Honneth, A. (2005). Autonomy, vulnerability, recognition and justice. In Christman, J. and Anderson, J. (eds.) *Autonomy and the Challenges to Liberalism: New Essays*. Cambridge: Cambridge University Press; 127–149.

Berlin, I. (1958). Two concepts of liberty. In Berlin, I. *Liberty*. Oxford: Oxford University Press (Hardy, H. ed.; 2002); 166–217

Bolton, D. and Banner, N. (2012). Does mental disorder involve loss of personal autonomy? In Radoilska, L. (ed.). *Autonomy and Mental Disorder*. Oxford: Oxford University Press.

Boyle, M. (2009). Active belief. *Canadian Journal of Philosophy*. Suppl. vol. 35: 119–147.

Buss, S. (2008). Personal autonomy. In E. N. Zalta (ed.) The Stanford Encyclopaedia of Philosophy (Winter 2008 Edition), URL = <http://plato.stanford.edu/entries/personal-autonomy/>.

Christman, J. (1989). *The Inner Citadel: Essays on Individual Autonomy.* New York: Oxford University Press.

Dunn, M. C., Clare, I.C.H. and Holland, A.J. (2008). 'To empower or to protect? Constructing the 'vulnerable adult' in English law and public policy'. *Legal Studies* 28(2): 234–253

Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.

Feinberg, J. (1986). *The Moral Limits of the Criminal Law: Vol. 3 Harm to Self*. New York: Oxford University Press

Fineman, M. A. (2008–2009). The vulnerable subject: anchoring equality in the human condition. *Yale Journal of Law and Feminism* 1(20): 1–23

Frankfurt, H. (1998). *The Importance of What We Care About*. Cambridge: Cambridge University Press.

Fulford, K.W.M. and Radoilska, L. (2012). Three challenges from delusion for theories of autonomy. In L. Radoilska (ed.) Autonomy and Mental Disorder. Oxford: Oxford University Press; 44–74.

Hart, H. L. A. (1963). *Law, Liberty, and Morality*. Oxford: Oxford University Press.

Heal, J. (2012). Mental disorder and the value(s) of 'autonomy'. In Radoilska, L. (ed.). *Autonomy and Mental Disorder*. Oxford: Oxford University Press.

Jaworska, A. (1999). Respecting the margins of agency: Alzheimer's patients and the capacity to value. *Philosophy and Public Affairs* 28(2): 105–138.

Kant, I. (1787). *Critique of Practical Reason* in *Practical Philosophy*, trans. Mary J. Gregor. Cambridge: Cambridge University Press, 1996, 5: 1–163.

Manson, N. and O'Neill, O. (2007) *Rethinking Informed Consent*. Cambridge: Cambridge University Press.

McLean, S. (2009). Live and let die. *British Medical Journal* 339:b4112.

Mill, J. S. (1859). *On Liberty*. Indianapolis (IN): Bobbs Merrill (1959).

Moran, R. (2001). Authority and estrangement: an essay on self-knowledge. Princeton: Princeton University Press.

O'Neill, O. (1984). Paternalism and partial autonomy. *Journal of Medical Ethics* 10: 173–178.

O'Neill, O. (2002) *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press

O'Neill, O. (2003). Autonomy: the Emperor's new clothes. *Proceedings of the Aristotelian Society*, Suppl. vol. 77: 1–21.

Oshana, M. (2006). *Personal Autonomy in Society*. Aldershot: Ashgate.

Plato (1992). *Republic*. 2nd ed. (transl. G.M.A. Grube). Indianapolis: Hackett Publishing Company.

Radden, J. (2012). Privacy and patient autonomy in mental healthcare. In Radoilska, L. (ed.). *Autonomy and Mental Disorder*. Oxford: Oxford University Press.

Radoilska, L. (2009). Public health ethics and liberalism. *Public Health Ethics* 2: 135–145.

Radoilska, L. (2012). Autonomy and Ulysses arrangements. In L. Radoilska (ed.), *Autonomy and Mental Disorder*. Oxford: Oxford University Press; 252–280.

Radoilska, L. (2013a). *Addiction and Weakness of Will*. Oxford: Oxford University Press

Radoilska, L. (2013b) Depression, decisional capacity, and personal autonomy. In K.W.M. Fulford, et al. (eds.) *The Oxford Handbook of Philosophy and Psychiatry*. Oxford: Oxford University Press; 1155–1170.

Shapiro, T. (2001). Three conceptions of action in moral theory. Noûs 35: 93-117.

Silver, M. 2002. Reflections on determining competency. *Bioethics* 16: 455 - 468.

Stoljar, N. (2000). Autonomy and the feminist intuition. In Mackenzie, C. and Stoljar, N. (eds.), *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*. New York: Oxford University Press; 94–111.

Taylor, J.S. (2005). Introduction. In Taylor, J.S. (ed.). Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy. Cambridge: Cambridge University Press; 1–29.

Tenenbaum, S. (2007). Appearances of the Good: an Essay on the Nature of Practical Reason. New York: Cambridge University Press.

Tenenbaum, S. (2009). Knowing the Good and knowing what one is doing. *Canadian Journal of Philosophy.* Suppl. vol. 35: 91–117.

van Willigenburg, T. and Delaere, P.J.J. (2005). Protecting autonomy as authenticity using Ulysses contracts. *Journal of Medicine and Philosophy* 30(4): 395–409.

Wiggins, D. (1995). Objective and subjective in ethics, with two postscripts about truth. *Ratio* 8: 243–258.