To appear in Ben Colburn, ed., *The Routledge Handbook of Autonomy*,
London: Routledge.

Autonomy and Mental Health

Hallvard Lillehammer

Abstract

The topic of mental health presents a distinctive set of problems for discussions of autonomy. First, there is a great deal of variation in the way autonomy is understood in the discussions of mental illness, disorder or incapacity. Second, on many of the standard ways of understanding autonomy it is at best unclear how the concept applies in the context of mental illness, disorder or incapacity. Third, with respect to ways of understanding autonomy that clearly do apply in the context of mental illness, disorder or incapacity it is at beast unclear why autonomy, so understood, should be thought of as being as important as it often is. In this chapter these problems are addressed by explicitly connecting questions about the value of autonomy to questions about what interests people have in virtue of their capacities to act and interact among ethically significant others.

The Challenge

The term 'autonomy' is frequently employed in discussions of mental health, both within and outside academia. Even a cursory survey of the contemporary literature is likely to reveal the following four things about this use. First, the term 'autonomy' is used in a number of different ways in the literature, not all of which are consistent with each other. Second, and even with respect to any given use of the term, it is not always clear exactly what 'autonomy' refers to in that use; nor that there is any single thing it does refer to. Third, whatever the term 'autonomy' refers to in the contemporary literature, it is usually assumed to be something very important. Fourth, whatever the importance assigned to autonomy in this literature, it is normally not assumed to be the only important thing (even though it is sometimes thought to be the most important thing). Faced with a complex theoretical landscape like this, it is only reasonable to hope for some clarification of the conceptual and theoretical landscape.

Autonomy has long been recognized as one of the basic values (or 'principles') of medical ethics, along with *best interest*, or what is sometimes labeled 'the principle of *beneficence*' (see e.g. Beauchamp & Childress, 2019, who further appeal to the principles of *non-maleficence* and *justice*). Considered as an issue of potential moral significance, an individual's autonomous judgment, wish or decision is something that anyone who cares for them is widely recognized as having an obligation to take account of. One way of doing so is to consider autonomy as a value to be weighted, or otherwise considered, along with other relevant values (such as the patient's psychological comfort). In this sense, the idea that autonomy is an ethically significant consideration is not a very controversial matter.

Another way of taking the value of autonomy into account is to regard it as a decisive limit or constraint on how to treat someone, or as something like an ethical 'trump card' that would normally defeat all other ethically relevant considerations at play in a given situation (see e.g. Dworkin 1984). Consider, for example, the case of a mental health patient who exercises their legal right to withdraw their previously given consent to a diagnostically recommended medical intervention. In this sense, the value of autonomy is a more controversial matter, especially in cases (of which there are many in the context of mental health) where the patient in question has been judged by others not to have a reasonable grasp of, or concern with, what is truly in their own best interest. Why should anyone think that people ought to have their decisions, judgment or wishes respected in matters regarding their own treatment in these and similar cases of likely or certain prudential failure or breakdown? This is one of the most serious challenges facing the application of autonomy-related considerations in the context of mental health. It is therefore on this challenge that much of the discussion in this chapter will focus.

The idea that individuals have a right to have their decisions, judgments or wishes respected on matters relating primarily to themselves is one of the most fundamental commitments of contemporary liberal societies, and one that has been a key element of 'Western' moral and political philosophy in some form at least since the Seventeenth Century (see e.g. Mill 1859; Schneewind 1997; Colburn 2010). According to one way of understanding this idea, an individual's right to autonomy is a crucial part of the freedom we owe to each other as the rational, self-governing agents that the liberal philosophical tradition has regarded many of us as being, or at least as capable of being on some occasions, or somehow in principle. (As the history of colonialism, gender relations, racism and ethnic strife ought to remind us, there has been considerable variation over time in exactly who this tradition has been prepared to include in this privileged class.) One central thought in this tradition is that it is precisely in virtue of our capacities as rationally self-governing agents that we can truly be thought of as empowered (whether by God or by Nature) to be the 'sovereign' masters of our own lives, and thereby also as capable of being held morally responsible for the (good or bad) consequences of our actions, as well as being treated as the proper subjects of corresponding rights and duties (see e.g. Dworkin 1994).

There have been plenty of calls for caution about this idea even within the liberal tradition; for example from those who think that the somewhat idealized conception of the rational, self-governing individual it has sometimes appealed to is based on a dubiously consistent or self-reinforcing fiction (see e.g. Geuss 2001). Yet even those who regard the paradigmatic versions of this idea with a certain degree of suspicion will nevertheless tend to agree with the following two claims, the joint plausibility of which are enough to frame the main challenge of this chapter. First, there are different degrees of self-governance on display among human beings in different situations; one undeniable example of impairment or failure of which is the case of severe mental illness, disorder or incapacity. Second, friends, family and medical professionals should normally be reluctant to coerce, manipulate or control their fellow human beings even in cases (such as severe mental illness, disorder or incapacity) where self-governance as imagined on the theoretical model just described is either in short or negligible supply.

The joint plausibility of these two claims gives rise to the following challenge. If mental illness involves individuals who fail to act in their best interest and fail to possess the standard features associated with self-governance; and if the failure of those individuals to act in their best interest is attributable partly to their failure to exhibit self-governance; then we need an alternative explanation of why it is reasonable to think that their judgments, decisions or wishes ought nevertheless to be respected; for example in extreme circumstances where lives, relationships or livelihoods are at stake (c.f. Feinberg 1986; Groll 2012; Lillehammer 2012; 2020; Radoilska (ed.), 2015). How, if at all, are we supposed to make sense of this presumptively reasonable attitude of respect for what in the contemporary literature is alternatively referred to as 'patient' or 'service user' autonomy'?

Mental Illness and the Absence of Autonomy

To make the discussion more concrete, it will help to consider the issue of autonomy and mental health in direct application to some recognized examples of mental illness, disorder or incapacity, and currently accepted conceptions thereof (c.f. Bolton 2008). The following four types of case have each been selected in order to illustrate the explanatory challenge articulated in the previous section; each type of case being in some way or other problematically related to one or more features of agency commonly associated with autonomy as self-governance.

i) Autonomy as self-governance is sometimes associated with the capacity for, or the exercise of, *rational agency* (see e.g. Kant 1785/1994; Korsgaard 2009). Yet some forms of mental illness are paradigmatically associated with the persistent failure or significant absence of rationality, including diachronic inconstancy; synchronic incoherence; self-deception; wishful thinking; lack of self-control; acting against one's better judgment; severe confusion; anxiety, and similar afflictions (see e.g. Pears 1984; Nussbaum 2009).

ii) Autonomy as self-governance is sometimes associated with the idea of the capacity for, or the exercise of, self-authorship or some other conceptually articulated *self-conception* on the part of the agent as a person living a certain kind of life (see e.g. Raz 1985; Velleman 2005). Yet some forms of mental illness, disorder or incapacity are paradigmatically associated with the loss or absence of any such capacity or exercise, including extreme cases of dementia or brain damage (see e.g. Buchanan 1988; Dworkin 1994; Shiffrin 2004).

iii) Autonomy as self-governance is sometimes associated with a set of ideas surrounding the capacity for, or the exercise of, a self-conception that is rooted in a generally *authentic, sound* or *truthful* grasp of basic facts about oneself, the world, and one's place in it (see e.g. Frankfurt 1971; Benson 1983; Oshana 2007). Yet some forms of mental illness, disorder or incapacity are paradigmatically associated with the reduction, loss or absence of any such capacity or exercise, including extreme or persisting forms of confusion, delusion or confabulation (see e.g. Bortolotti et. al. 2010; c.f. Martin 2007).

iv) Autonomy as self-governance is sometimes associated with the capacity for, or the exercise of, actions, plans and decisions that are *independent* of the authority, power or non-rational influence of others (see e.g. Mill 1859/1989; Dworkin 1998). Yet some forms of mental illness are paradigmatically associated with being highly dependent on the assistance, intervention or decision-making authority, power or non-rational influence of others; including some, even if not all, the conditions described in i)-iii) above.

In the case of each of the categories i)-iv), there is a real sense in which the presence of mental illness, disorder or incapacity is strongly correlated with the failure to display at least some of the standardly recognized marks of autonomy as self-governance. The challenge of making sense of the value of autonomy in the context of mental health is therefore in large part the challenge of working out what to say about appeals to the value of autonomy in cases that fall into one, or more, of the categories i)-iv) above.

In other to make progress on this challenge, we should bear in mind the following two possibilities. The first is that what justifies taking account of a mentally ill, disordered or incapacitated agent's decisions, judgments or wishes in any individual case could in principle have little or nothing to do with respecting the value of autonomy, however that notion is best understood. For example, in some cases the most important consideration speaking in favor of respecting someone's decisions, judgments or wishes might be the desirability of avoiding discomfort or frustration on the part of that someone, or the desirability of finding some practically feasible way to address their persistent and unreasonable refusal to cooperate. The second is that insofar as what justifies taking account of a mentally ill, disordered or incapacitated person's decisions, judgments or wishes does have anything to with the value of autonomy, this need not be the same thing in every case because what is meant by 'autonomy' in one context of decision is very different from what is meant by it in another (c.f. Arpaly 2002; Heal 2012). In the discussion that follows it is important to bear in mind that each of the arguments to be considered needs to be evaluated against the background of these possibilities.

Before considering some of the harder challenges presented by cases of mental illness, disorder or incapacity, it is also important to note that in many such cases there is no serious obstacle to understanding how considerations of autonomy as self-governance do apply. First, it is relatively uncontroversial that many cases of mental illness or disorder do not seriously remove an agent's capacity for self-governance, at least not permanently. Thus, if I am going through a period of depression, that does not necessarily remove (although it might well disturb) my basic ability to think of myself as a certain kind of person who would quite like to live a less troubled kind of life, or to make strategic judgments about how I ought to handle my depressive episode as a result. Indeed, my basic capacity to consider the diagnostic situation in this way (possibly with the help of others) could be the very thing that makes me seek assistance, therapy, intervention or some other relevant interaction in the first place (see e.g. Rashed 2019).

Second, we do well to remember that even in cases where mental illness, disorder or incapacity do conflict with a patient's capacity for self-governance, the value of autonomy understood in terms of self-governance can still be significantly operative in care or treatment insofar as it is a value that either orients or otherwise sets the terms in which care, treatment or intervention is proposed (see e.g. Pickard 2015). Among the wide range of cases in which considerations of autonomy as self-governance can play an aspirational role in this way are: i) *proleptic* appeals to self-governance (e.g. where treating someone as a self-governing agent is used as a means to enhance their capacity for self-governance); ii) *preventive* appeals to self-governance (e.g. where treating someone as a self-governing agent is used as a means to prevent or slow down their loss of capacity for self-governance); iii) *symbolic* appeals to self-governance (e.g. where treating someone as a self-governing agent is used to express recognition of the fact that they once were a self-governing agent and remain to be counted as such the purposes of making decisions that primarily affect themselves); or iv) *statistical* appeals to self-governance (e.g.

where treating someone as a self-governing agent is regarded as a precautionary or statistically justified norm in a context where the precise capacities of every individual so treated is regarded as an uncertain or a insignificant 'outlier'). Although this list is not exhaustive and some items on it are likely to be more controversial than others, it does serve to illustrate the basic coherence of a widely recognized but easily forgotten thought. This is that it sometimes makes sense to regard a person as being a certain way (in this case a substantially self-governing agent) in order to locate them within a given ethical category, even if it is at best an open question whether they do in fact, at that time, literally possess all the features in terms of which that category is paradigmatically defined. (Consider, by analogy, official government advice that a large group of people should 'self-isolate' during the course of a pandemic, in spite of the fact that it is known that a large proportion of the individuals in that group are actually perfectly healthy or very unlikely to be badly affected by the disease.)

Mental Illness and Autonomy Beyond Self Governance

In some of the cases discussed in the previous section, medical professionals in many jurisdictions would normally appeal to a legally recognized 'fallback' condition of *capacity* or *competence* in order to decide whether or not to constrain their interventions by the decisions, judgments or wishes of patients who fail to display one or more of the standard marks of autonomy as self-governance (see e.g. Mental Capacity Act 2005; Alzheimer Society 2018). It might therefore be thought that it is not the somewhat idealistic or aspirational value of autonomy as self-governance that should be thought of as operative in these and similar cases of medical care, treatment or intervention, but rather the more down-to-earth value of respecting a patient's basic capacity or competence; where capacity or competence could be understood in terms of the idea of someone being able to do no more than: i) *understand* the information relevant to a decision; ii) *retain* that information for a contextually specified period during which the decision is being considered; iii) *use* or *weigh* that information in making a decision; and iv) *communicate* a decision either verbally or in some otherwise intelligible manner.

It is, indeed, an indispensable historical insight that capacity or competence has widely been recognized as the legally decisive fallback condition on valid patient consent in cases of medical intervention where other more demanding aspects of autonomy as self-governance are temporarily or permanently absent. What to make of this fact from a critical perspective is, however, a different matter. Perhaps it might be argued on this basis that the idea of autonomy as self-governance is not in fact as central to the ethics of mental health as many philosophers and other commentators have tended to assume. Or perhaps it might be argued on this basis that when autonomy is invoked in discussions of mental health, it is really neither more nor less than capacity or competence that people ought to be charitably interpreted as having in mind; even if from the more abstract and theoretical perspective of moral and political philosophy this use of the term 'autonomy' is at best misleading and at worst inadvisable.

Each of these alternative hypotheses has a genuine claim to our serious attention. Yet neither hypothesis will suffice to fully respond to the main challenge of this chapter. That this is so becomes clear once we recall what actually needs to be explained. For even if we agree that the question of capacity or competence is strongly correlated with the question of whether something is a serious ethical consideration that genuinely merits our respect, we still don't have an answer to the question of why we should focus our ethical attention on someone's capacity or competence in the absence of a more substantial capacity for self-governance; for example to the extent of thinking that capacity or competence on its own is capable of generating 'trump-like' constraints against beneficent medical or other paternalistic interventions. This question is especially acute in cases (of which there are many in the context of mental health) where the decisions made by the patient who is deemed to have basic capacity or competence are judged by friends, family or medical professionals as not being in that patient's own best interest. The case of someone with anorexia nervosa who refuses food or other forms of nourishment; the case of a person with advancing dementia who refuses the protections from their own future actions promised by safe and effective social care; or the case a person whose decision making is regularly affected by persistent and irrational delusions each fall into this class of 'non-ideal' cases in which the rationale for respecting patient 'autonomy', however understood, remains to be fully accounted for (c.f. Bortolotti et. al. 2012).

In any case, the appeal to capacity or competence – whatever its ultimate justification – is unable to capture the full range of mental illness, disorder or incapacity in which the question of patient autonomy has historically been thought to arise. Common approaches and attitudes towards mental illness, disorder or incapacity include norms against interference with the decisions, judgments or wishes of others, where those norms are often regarded as applying well beyond the domain of capacity or competence. Among norms of this kind that are widely observed both within and outside the domain of clinical practice is a basic reluctance to impose oneself on others during the course of a vast range of everyday interactions, such feeding, washing, exercise or play. That patients deemed to have capacity or competence are not the only ethically significant agents who need to eat, wash, or keep their minds and bodies in a basically functional condition is a truth so obvious it ought not to require stating. Yet the attractions of an approach to autonomy and mental health that rests by postulating a threshold of capacity of competence for the duty to respect the decisions, judgments or wishes of mental health patients or service users are quickly reduced once we consider the existence of this additional group of cases to which it fails to speak.

Perhaps it is for cases within this 'non-ideal' class of agency that it is most tempting to invoke some minimally demanding idea of individual autonomy, such as one that requires no more of agents to be autonomous than that they are in a position to make or have decisions, judgments or wishes, which can then be respected in a purely 'negative' sense. Thus, there is a widely recognized use of the term 'autonomy' both in medical ethics and elsewhere in which that term implies nothing over and beyond *not being coerced, manipulated or otherwise interfered with* in relation to decisions, judgments or wishes one is disposed to arrive at 'independently', or 'for oneself' (c.f. Miller & Wertheimer, 2010). If we take this minimally demanding use of the term at face value, perhaps we shall have no residual problem in accounting for how people (or, indeed, some other sentient creatures) can have their autonomy respected even in the kind of 'non-ideal' cases where neither autonomy as self-governance nor capacity/competence can plausibly be said to apply.

What we should ask about such minimally demanding uses of 'autonomy', however, is what explains why people think they *ought to value* autonomy thus understood, for example to the point of considering it an ethically important enough consideration to 'trump' other ethically relevant considerations in a given situation; including what friends, family or medical professionals consider to be in the patient's own best interest. The problem, in a nutshell, is that the moral or political appeal to independence or non-interference cries out for a backup story that links independence or non-interference to some valuable feature on the part of their subjects or their interpersonal relations that explains why we should care so much about autonomy understood in this primarily 'negative' way. In standard cases of coercion, interference or intervention discussed in the professional literature it is often exactly the value of autonomy considered as self-governance that is said to provide that explanation. Yet in the absence of autonomy as self-governance, it is incumbent on those who insist on respecting the decisions, judgments or wishes of patients who suffer from some form of mental illness, disorder or incapacity to offer some alternative explanation of its moral or political value. To put the challenge in a different way: why should we worry about coercing or otherwise acting against the consent of a pre-, post-, or otherwise not self-governing agent in cases where their ‘right to go wrong’ with respect to what is in their own best interest cannot be justified by appeal to their status as self-governing, or otherwise ‘sovereign’, individuals? To this question we now turn.

Mental Illness and Agency-based Interests

Although some philosophical discussions of mental health understand autonomy and interests as two separate factors in determining priorities for care, treatment or intervention it would be wrong to think of these factors as always independent or competing. This is not only because in some cases agents who are granted the freedom to choose will actually decide in favor of what is in their own interest. It is also because some of the things that can plausibly be considered as being in someone’s interest are the very kinds of thing that are sometimes appealed to under the heading of ‘Autonomy’. For example, being granted a significant range of freedom from interference; being granted substantial independence in decision-making; or being meaningfully included in decisions that affect oneself can each be considered as being in one’s interest in the sense that one’s life would be judged as going better or worse depending on the extent to which one is granted the relevant freedom or participation in question. In other words, one of the things that can reasonably be thought to respect the interests of a mentally ill, disordered or incapacitated agent is not removing their opportunities to exercise their agency to whatever extent their condition allows.

Some of these ways of exercising agency may involve capacities that have traditionally been associated with autonomy understood as self-governance. Yet not all of them are. And some of the things that are not associated with autonomy as self-governance involve agential capacities that are available to individuals whose actual condition is very different from that of the somewhat idealized agents that philosophers have sometimes had in mind when using the language of autonomy as self-governance. Perhaps, therefore, by thinking about our basic challenge in terms of these recognizable, but less demanding, ways of exercising agency we can expand the range of cases in which we are able to explain the rationale for norms that speak in favor of respecting the decisions, judgments or wishes of ethically significant others so as to include the kinds of 'non-ideal' cases to which philosophical appeals to the value of autonomy have sometimes failed to speak. In other words, by appealing to a set of variously realized 'agency-based' interests, we might be able to explain what is wrong about treating mentally ill, disordered or incapacitated agents in ways that make them either i) excluded from taking part in decisions that involve themselves; ii) have their own wishes systematically frustrated by well-meaning others; iii) have such forms of agency as they are capable of ignored, undermined or destroyed; or iv) not be given a voice of their own in what is inevitably a highly asymmetric power relationship with the friends, family, or medical professionals who care for them (c.f. Kittay 2015; c.f. MacKenzie & Stoljar 2000).

The interest in playing a part in decisions that affect us is one that extends far beyond cases of mental illness, disorder or incapacity. Indeed, it is an interest that is observable across a wide range of interactions where not being so included is a cause of complaint on the basis of unwarranted exclusion. To take one trivial example, implicit recognition of this interest is one of the reasons why appeals to *hypothetical consent* are frequently regarded as not being fully legitimating on their own, even in cases where the not actually consenting party is undeniably in a less than a rationally excellent condition compared with those who presume to rely on that hypothetical consent. Consider, for example, a case where you reasonably consider me irrationally suspicious about your ability to pay me back any money you owe me, with interest, in full, and on time. It would normally raise an eyebrow or two if you therefore went ahead and just took the money, giving as your excuse that I would have lent you the money if I had rationally considered all the available evidence as to why there is really no serious possibility that you will ever default. This response is only one of many similar cases in which a deeply felt concern about social, institutional or deliberative exclusion manifests itself when agents are sidelined, silenced or alienated from decisions that substantially affect themselves and their relations to others.

If we are entitled to think that these and similar interests in being actually included in decisions that affect ourselves are sufficiently important to serve as a justification for the imposition on others of relevant *duties* (such as a duty to actually ask and then show due respect for the answer), there is a *prima facie* case for saying that those interests are also grounds for the attribution of corresponding *rights,* in precisely the sense of the metaphorical 'trump cards' alluded at the start of this chapter. The question whether this is the most plausible way to think about the relationship between autonomy and interests on the one hand, and rights and duties on the other, is a matter of controversy. Although some aspects of that controversy relates to matters that go beyond the remit of this chapter (see e.g. Raz 1986), other aspects are directly related to the core ideas that have been at issue in the previous sections, including the question of why we should think of the agency-based interests at issue in cases of mental illness, disorder or incapacity as sufficient to justify the attribution of the relevant rights and duties in the absence of a back-up explanation framed either in terms of autonomy as self-governance, or in terms of capacity or competence.

There is more than one reason to be cautiously optimistic on behalf of the language of agency-based interests when it comes to addressing this question, even in the 'non-ideal', or hard, cases that have been the focus of discussion in this chapter. Here I shall mention two.

First, and most importantly, many of us who have a stake in these issues are able to recognize the force of the concern that our treatment of significant others is ethically problematic to the extent that the agency of those others is excessively *objectified*, in the sense that instead of being regarded as another thinking, feeling and acting subject with a unique perspective on the world, their thoughts or wishes are treated as if they were ideally reducible to the reliably manipulated outputs of a physical mechanism. When responding to the agential resistance of an ethically significant other what we are doing amounts to more than dealing with a purely material obstacle to the promotion of some independently specifiable good (such as measurable contentment, or institutionally predictable and manageable behavior).

Far from being unique to our relationship with rationally self-governing human adults, this concern about excessive objectification is widely observable in relationships between human beings quite independently of questions about their capacity or competence, and in relationships between human beings and other sentient creatures as well.

Second, it does not follow from the claim that invocations of rights have a 'trump' like function that the rights so invoked are absolute. On the contrary, this cannot possibly be the case insofar as there could always be more than one right at work in a given situation, and one of which (e.g. a right to life) might reasonably be thought of as more important than another (e.g. a right to free movement). Moreover, some relatively trivial rights (such as a right to have a say on some trivial matter) could potentially be overridden by an appeal to important values (such as the value of maintaining an hygienic clinical environment). Hence, it does not follow from the claim that agency-based interests are important enough to constitute the basis of rights that agency-based interests are always *the most important*, as opposed to *one very important*, thing in the circumstances. What does follow is that whenever agency-based interests are ignored or overridden, those who ignore or override them have something to answer for (e.g. by way of apology, explanation, restoration or some alternative form of adjustment).

In closing, it must be emphasized that to describe the ethical significance of the agency-based interests of severely disabled or incapacitated agent in the language of rights and autonomy is not in any way intended to romanticize, trivialize, or otherwise downplay the extent of the extraordinary suffering and tragedy that is often involved in living, living with, or otherwise managing the life of someone suffering from a severe case of severe mental illness, disorder, or incapacity. On the contrary, to insist on this point points is arguably one of the prerequisites of understanding the ethical significance of that suffering or tragedy, and the combination of concern and uncertainty that people often experience with respect to the care and treatment of those who are the primary victims of it.

Related Topics

The reception and afterlife of Kantian autonomy; Rationality; Self-authorship; Social relations; Autonomy and responsibility; Coercion and autonomy; Respect for autonomy; Autonomy and well-being; Autonomy and rights; Autonomy and consent; Autonomy in dying and death; Autonomy and paternalism; Autonomy in the law; Autonomy and consent in clinical and caring practice.

References

Alzheimer’s Society GB (2018) “Decision-making and Mental Capacity,” Online: https://www.alzheimers.org.uk/get-support/legal-financial/decision-making-and-mental-capacity#content-start (Accessed July 25 2018).

Arpaly, Nomy (2002) *Unprincipled Virtue: An Inquiry Into Moral Agency*, Oxford: Oxford University Press.

Benson, John (1983) “Who Is the Autonomous Man?,” *Philosophy* 58: 5-17.

Beauchamp, Tom L. & Childress, James F. (2019) *Principles of Biomedical Ethics*, Eighth Edition, Oxford: Oxford University Press.

Bolton, Derek (2008) *What is Mental Disorder? An Essay in Philosophy, Science and Values*, Oxford: Oxford University Press.

Bortolotti, Lisa, Cox; Rochelle; Broome, Matthew & Mameli, Matteo (2012) “Rationality and Self-knowledge in Delusion and Confabulation: Implications for Autonomy as Self-governance,” in Lubomira Radoilska (ed.) *Autonomy and Mental Disorder*, Oxford: Oxford University Press, pp. 100-122.

Buchanan, Allen (1988) “Advance Directives and the Personal Identity Problem,” *Philosophy and Public Affairs* 17: 277-302.

Colburn, Ben (2010) *Autonomy and Liberalism*, London: Routledge.

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

Dworkin, Ronald (1984) ‘Rights as Trumps’, in Jeremy Waldron (ed.) *Theories of Rights*, Oxford: Oxford University Press, pp. 153-67.

Dworkin, Ronald (1994) *Life’s Dominion*, New York: Vintage.

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

Frankfurt, Harry (1971) “Freedom of the Will and the Concept of a Person,” *The Journal of Philosophy* 68: 5-20.

Geuss, Raymond (2001) *History and Illusion in Politics*, Cambridge: Cambridge University Press.

Groll, Daniel (2012) “Paternalism, Respect, and the Will,” *Ethics* 122: 692-720.

Heal, Jane (2012) “Mental Disorder and the Value(s) of “Autonomy””, in Lubomira Radoilska (ed.) *Autonomy and Mental Disorder*, Oxford: Oxford University Press, pp. 3-25.

Kant, Immanuel (1785/1994) *Groundwork of the Metaphysics of Morals*, Cambridge: Cambridge University Press.

Kittay, Eva, F. (2005) “At the Margins of Personhood,” *Ethics* 116, 100-131.

Korsgaard, Christine (2009) *Self Constitution*, Oxford University Press.

Lillehammer, Hallvard (2012) “Autonomy, Value and The First Person,” in Lubomira Radoilska (ed.) *Autonomy and Mental Disorder*, Oxford: Oxford University Press, pp. 192-213.

Lillehammer, Hallvard (2020) “Autonomy, Consent and the Non-Ideal Case,” *The Journal of Medicine and Philosophy* 45: 297-311.

MacKenzie, Catriona & Stoljar, Natalie, eds. (2000) *Relational Autonomy: Feminist Essays on Autonomy*, Oxford: Oxford University Press.

Martin, Adrienne M. (2007) “Tales Publicly Allowed: Competence, Capacity, and Religious Belief,” *The Hastings Center Report* 37: 33-40.

Mental Capacity Act (2005), Legistation.gov.uk Online, http://www.legislation.gov.uk/ukpga/2005/9/contents (Accessed 6 April 2020).

Mill, John Stuart (1859/1989), *On Liberty and Other Writings*, Cambridge: Cambridge University Press.

Miller, Franklin G. & Wertheimer, Allen (eds.), (2010) *The Ethics of Consent*, Oxford: Oxford University Press.

Nussbaum, Martha (2009) "The Capabilities of People with Cognitive Disabilities," *Metaphilosophy* 40: 331-351.

Oshana, Marina (2007) "Autonomy and the Question of Authenticity," *Social Theory and Practice* 33: 411-429.

Pears, David (1984) *Motivated Irrationality*, Oxford: Oxford University Press.

Pickard, Hanna (2013) "Responsibility without Blame: Philosophical Reflections on Clinical Practice," in K. W. M. (Bill) Fulford et. al. (eds.) *The Oxford Handbook of Philosophy and Psychiatry*, Oxford: Oxford University Press, pp. 1134-1154.

Radoilska, Lubomira, (ed.), (2012), *Autonomy and Mental Disorder*, Oxford: Oxford University Press.

Radoilska, Lubomira (2015) "Autonomy in Psychiatric Ethics," in John Z. Sadler (ed.) *The Oxford Handbook of Psychiatric Ethics*, Oxford: Oxford University Press, pp. 354-371.

Rashed, Mohammed Abouelleil (2019) *Madness and the Demand for Recognition: A Philosophical Inquiry into Identity and Mental Health Activism*, Oxford: Oxford University Press.

Raz, Joseph (1986) *The Morality of Freedom*, Oxford: Oxford University Press.

Schneewind, Jerome (1997) *The Invention of Autonomy*, Cambridge: Cambridge University Press.

Shiffrin, Seana (2004) “Autonomy, Beneficence and the Permanently Demented,” in Justine Burley (ed.) *Dworkin and His Critics*, Oxford: Blackwell, pp. 195-217.

Velleman, J. David (2005) *Self to Self*, Cambridge: Cambridge University Press.