

[Forthcoming in L. Radoilska (ed.), *Autonomy and Mental Disorder*, OUP, 2012]

Autonomy, Value and the First Person

Hallvard Lillehammer

1. Introduction

This paper has three parts. In the first part I say something about how ethical questions concerning autonomy look different when approached from a first-person, as opposed to an other-person, perspective on practical reasoning. In the second part I distinguish between two different kinds of autonomy that might be at stake in discussions of ethics and mental health. In the third part, I say something about how ‘autonomy’ so understood relates to value and the promotion of desirable outcomes. In each case, my guiding question is what it could mean to respect autonomy in cases where doing so is not conducive to the promotion of desirable outcomes on account of the fact that the choices of the person whose autonomy is respected are affected by some kind of mental disorder or incapacity.

2. Autonomy and the first person

Considerations of autonomy normally play a different role in ethical thought depending

on whether we take a first-person or an other-person perspective (by which for present purposes I include both a second-person and a third-person perspective) on practical reasoning (c.f. Darwall 2006). For many of us and for much of the time, at least when things go well, autonomy is what I call ‘reflectively transparent’ from the first-person point of view. What I mean by this is that considerations of autonomy very often play no substantial role in the content of our practical reasoning, which when things go well is directly targeted at the pursuit of what we see as desirable outcomes. Thus, the thought that my efforts to settle on ends and to take means to those ends had better not be coercively interfered with by others does not normally play a part in my practical reasoning when I wonder what product on the shelf I should go for at Boots while shopping for toothpaste. Instead I focus directly on getting the best toothpaste, given what I take my needs and preferences to be. In so doing, I implicitly take my autonomy thus understood for granted. Similarly, the thought that I might not be a very good judge of what is best for me all things considered does not prevent me from making an effective decision when I choose between Colgate and Boots’ own brand. Perhaps I read the information provided. Perhaps I ask the pharmacist for advice. But in the end I will normally assume that the stage is set for me to make a final decision based on my own judgement. The thought that in buying some toothpaste I am exercising my autonomy as a self-governing agent does not enter my practical reasoning in this context. I simply aim to get some decent toothpaste. Indeed, the introduction of explicit thoughts about the ethical significance of my autonomy would as likely as not be counterproductive, and in the worst case would short-circuit my attempt to get some decent and affordable toothpaste by leaving me flustered and uncertain, unable to make what is a comparatively

simple and trivial decision.

This presumption of reflective transparency often disappears when we move to an other-person perspective on agency. It is a well-known fact that we constantly disagree about which options are best to take, even on matters as simple and trivial as buying toothpaste. Suppose, for example, that I am about to buy the most expensive toothpaste in the shop. In your view, there is no relevant difference between this toothpaste and Boots' own brand, the latter being much cheaper and tasting much the same. It is simply a matter of branding. If so, thoughts about autonomy may enter your practical reasoning from a second-person perspective to the extent that although you may fairly try to convince me that I am wasting my money in buying the more expensive label, you will normally stop short of coercing me into not doing so. In this way, considerations of autonomy sometimes figure in the foreground of practical reasoning from an other-person perspective in a way they do not from a first-person perspective. To respect someone's autonomy in this sense is in part to respect their freedom to make a mistaken (or at least a sub-optimal) decision. That is what you are granting me by not coercively interfering, and the thought of doing so is in no way paradoxical from your other-person perspective. From my first-person perspective, on the other hand, the idea that you are respecting my autonomy to make a mistaken (or sub-optimal) decision will not normally play any direct role in my deliberation there and then without appearing paradoxical. Unless I have either conferred on you the authority to decide for me or am being deliberately perverse, I will simply choose what I judge to be (in some relevant sense) the best option. The fact that I may also, from some abstract or theoretical point of view, value the freedom to make my

own mistakes no more undermines my claim to be making a sensible decision in this individual case than the fact that some scientific authors apologise in advance for mistakes in their works thereby undermines their claim to believe every individual sentence put forward in those works.

When considerations of autonomy enter practical reasoning from the first-person perspective, this is often because the agent's own capacity to choose well is in question. First, someone might try to take my freedom over some decision away from me or prevent me from having it in the first place. Legal constraints work that way, as do various forms of personal and social oppression. It is natural for us to protest at restrictions on our freedom when we judge them unjustified, and it is natural to appeal to some conception of autonomy in doing so. Second, there are some options about which we may know that we are not good, or even competent, judges. Thus, there are products that, in virtue of their nature and use, I do not consider myself well placed to judge the value of. If I need a drug to cure a serious intestinal infection I will normally be willing to grant authority over its selection to a relevant health professional and will normally judge that I ought to do so.¹ Third, there are situations in which I may doubt whether I am a good judge about an arbitrarily wide range of questions relating to my own best interest. If I am going through a period when I am beset by constant panic attacks I may consider having the power of decision about a range of issues taken away from me and handed to

¹ I may still, of course, expect my autonomy to be respected by having the pros and cons of different drugs explained to me and by not having any drug forcibly administered against my will.

another person. In doing so, I may actually be placing a high value on autonomy in some sense even as I hand over control of some aspect of my life to someone else. If I present myself voluntarily for extended psychiatric treatment, for example, I might be placing myself in this category. A more difficult case is where my capacity to choose well is put into question and I either mistakenly refuse to admit that my ability to choose well is seriously compromised, or where I refuse to hand over control of some aspect of my own life in spite of having previously judged correctly that my ability to choose well is seriously compromised.² In such cases, it may be tempting to describe the situation as one in which considerations of autonomy either might, or ought to, have figured centrally in practical reasoning from the first-person perspective, but did not (or not in the right way). It may also be tempting to think that in such cases the absence of considerations of autonomy from the first-person perspective combined with the absence of a capacity to judge well implies that the normal considerations associated with autonomy as an ethically significant value are not in play: the only (or main) duty held by others towards me is to act in what is my best interest. In other words, for persons incapable of exercising their autonomy the question of respecting their autonomy need (or should) not arise.

This kind of scenario gives rise to a puzzle about our ethical relationship to mentally ill or

² The first case may capture an aspect of my personality if I am prone to exhibit certain forms of manic behaviour. The second case may capture an aspect of my personality if I am suffering from some form of mental incapacity, such as advancing dementia.

other psychologically challenged individuals who are capable of forming, expressing and acting on judgements about what is in their best interest, but who systematically fail to either judge well about their own self-interest or to successfully translate such judgements (well-founded or otherwise) into prudentially coherent plans or actions. On the one hand, such patients are sometimes described as incapable of judging, choosing, or acting autonomously to the point that even their personhood is put into question (c.f. McMahan 1996; Kittay 2005). So it might seem permissible to act towards them in the way we judge to be their own best interest from an other-personal perspective without giving any thought to considerations of autonomy. On the other hand, such patients are human beings towards whom it is natural to think owe a kind of respect that extends beyond the care we extend to other beings with a bare capacity for agency or sentience (Kittay 2005; Nussbaum 2009). So it might seem impermissible to treat them in whichever way we reasonably judge is in their own best interest regardless of how the situation appears to them from their own first-person perspective on practical reasoning. The first thought may seem to be an inescapable implication of the facts of the case. The second thought may seem to be an inescapable part of any ethics that is distinctively human (c.f. Williams 2006; Singer 2009). When combined, these two thoughts may reasonably be thought to be in serious tension. One aim of this paper is to get clearer about what this tension amounts to. A second aim is to expose one source of danger embodied in the move towards an other-personal perspective on agency in cases where things are not going well, namely the danger of losing sight of the fact that even individuals who suffer from mental illness or incapacity have a first-personal perspective on agency that can matter from an ethical point of view.

3. Two kinds of autonomy

The term ‘autonomy’ is used in a wide variety of ways. This fact is a cause of much disagreement and confusion. In this paper I make use of two distinctive interpretations of ‘autonomy’, each of which has had some currency in the philosophical and medical literature. I do not claim that these two interpretations are the only coherent interpretations of the term ‘autonomy’ in these contexts, or that other interpretations of the term either can or ought to be analysed in terms of these two. I believe the appropriate attitude to take towards the many different interpretations of ‘autonomy’ in medical ethics and elsewhere is one of pluralism. I shall not, however, defend this claim here (c.f. Christman 1988; Arpaly 2003; Beauchamp 2005).

On my first interpretation, ‘being autonomous’ means not being subject to coercive interference. Autonomy thus understood is a state of someone (or something) being (negatively) free, or independent, with respect to a certain range of options. Thus, I can be autonomous in this sense with respect to whether or not to take up residence for tax purposes in Spain, England, or Norway, even if I am not free in the relevant sense to either take up residence for tax purposes somewhere or not to take up such residence at all. I shall henceforth refer to autonomy so understood as ‘choice autonomy’. The idea of choice autonomy has played a central role in contemporary debates in medical ethics (c.f. O’Neill 2002). Respect for choice autonomy is embodied in the idea that health professionals normally have a duty to acquire informed consent for medical procedures

and treatment plans, that patients normally have a right to refuse treatment (even after having given informed consent), and that the normal relationship between carer and patient is one in which everyday interactions, such as taking meals or going to the toilet, are constrained by the need to respect the patient's wishes where this is practical. There has been much discussion among philosophers about what does, and what does not, qualify as coercive interference and voluntary choice (c.f. Nozick 1997). I shall have nothing new to say about this here. Instead, I stipulate that, for present purposes, an agent has choice autonomy with respect to a given option *only if* the agent is faced with a non-empty set of (in the relevant circumstances) reasonable options between which to choose (c.f. Raz 1986). I take this to exclude from the domain of choice autonomy subjection to medical treatment as a result of physical force, psychological manipulation, serious and credible threats, and the like.

One central question in the ethics of medical ethics is whether, and if so when, it is ever reasonable to deny someone choice autonomy in the sense just defined. Thus, it is sometimes argued that it is justifiable to subject human beings who suffer from severe mental illness or disability to coercive medical treatment in order to reduce the risk of harm to themselves or others, even if incurring the risk of such harm is not taken to licence the coercion of other persons in a normal (but often less than perfect) state of mental health or intellectual ability (c.f. Wikler 1979; Scoccia 1990). Given a prior ethical commitment to respect choice autonomy, such a distinction between when it is, and when it is not, justified to subject someone to coercive medical treatment stands in need of explanation.

On my second interpretation, ‘being autonomous’ means being a self-governing agent. Autonomy so understood is a state of realising a set of higher-order capacities of rational thought and agency, whereby practical options are reflectively endorsed and plans of action formulated or brought to execution (with or without the assistance of others). There is a wide range of conflicting accounts of what does, and what does not, qualify as genuine self-governance (c.f. Frankfurt 1986; Dworkin 1988; Korsgaard 2009). I shall make no attempt to choose between these accounts here. Instead, I stipulate that any plausible account of autonomy in this sense will include the following four conditions as necessary for genuine self-governance. First, self-governance requires the actual manifestation of a capacity of higher order reflection and endorsement of practical options (where under ‘practical option’ I include the option to have one set of motivations rather than another and ‘some, non-trivial’ is meant to exclude a purely latent potentiality but not to require anything like a ‘maximal’ realisation of this capacity). Second, self-governance requires the actual manifestation of a capacity for planning and executing actions that accord with practical options endorsed (where this is taken to imply a non-trivial degree of practical consistency and the formulation and execution of plans over time). Third, self-governance requires that an agent’s reflection, endorsement and execution of practical options are each responsive to minimally intelligible standards of rational argument (where this is taken to exclude extreme forms of irrationality but to include the scope for substantial disagreement about what does, or does not, count as good or bad reasons for and against a practical option). Fourth, self-governance requires reflection, endorsement and execution of practical options that involve a conception of

oneself as a single person living a certain kind of life (i.e. having a substantial self-conception), although the conception in question need not include in its scope one's entire life as a whole or be in any sense maximally unified (c.f. Dworkin 1994). I shall henceforth refer to autonomy so understood as 'agent autonomy'.

One central question in the ethics of mental health is whether, and if so when, it is reasonable to appeal to agent autonomy when deciding how to respond to someone's expressed or assumed preferences. Thus, it might be asked whether the informed consent of a mentally ill, but legally competent, patient is ever sufficient to establish the ethical permissibility of treating them in the way consented to, given that the informed consent of a mentally ill patient may well conflict with the judgement they would make if they were (in some suitable sense) genuinely self-governing, and therefore agent autonomous.³ Given a prior commitment to respect agent autonomy, any decision to either include or exclude appeals to agent autonomy in treatment decisions where the patient is thought to not meet the conditions of agent autonomy stands in need of explanation.

³ The current legal understanding of competence in the United Kingdom classifies a patient as having capacity or competence if they i) can understand and retain information relevant to the decision in question, ii) believes it, and can reflect on that information to arrive at a choice and use that information as part of the decision-making process, and iii) can express or otherwise communicate that choice (Alzheimer's Society 2010). It is widely agreed that many persons diagnosed with some form of mental illness can satisfy each of these criteria at the same time. For further discussion of the legal definition of mental incapacity, see Holroyd [this volume].

The idea of agent autonomy is a core component of the liberal individualism defended in much contemporary moral and political philosophy. In particular, the idea of agent autonomy plays a central role in some very influential accounts of the individual as a potential duty holder, and thereby also a potential holder of the rights that correspond with such duties (c.f. O'Neill 2000).⁴ First, a genuinely self-governing individual is a paradigm candidate for being someone who has the capacity to make, be held to, or hold others to, ethically or legally binding promises or contracts the ethical significance of which they can themselves understand. Second, the self-governing individual is someone who has the capacity to form and execute their own reflectively endorsed life-plans, and to whom reasons are sometimes said to be 'owed' by others (whether individuals or institutions) to justify actions that interfere with or otherwise undermine those life-plans (c.f. Scanlon 1998). Third, the self-governing individual is someone who has the capacity to make ethically significant requests or decisions about what should happen to them in conditions of extreme vulnerability, such as a demented state of life beyond reason (c.f. Dworkin 1994). Fourth, the self-governing individual is the standard model of the liberal citizen, whose capacity for autonomous agency is often said to stand as a normative limit on the promotion of impartial and other social goods (c.f. Rawls 1970). On this view, the self-governing individual is someone who has the capacity, and therefore the right, to

⁴ Some philosophers argue that the domain of right holders extends beyond the domain of duty holders, rights being identified with special justifications for imposing a certain duty (c.f. Raz 1986). I return to the relevance of this debate below in my discussion of duties of care to non-autonomous agents.

participate in an arbitrarily wide range of mutually beneficial but inherently risky cooperative practices, such as market exchanges or the democratic process. In these and other ways, some idea of autonomy as self-governance has played a central role in liberal explanations of the idea that there is a distinctive kind of respect that we owe to each other as moral agents, and that the rationale for this respect is intimately connected to our capacity for self-conscious reflection and endorsement of ethically significant goals.⁵

Choice autonomy is possible in the absence of agent autonomy. Some agents are incapable of agent autonomy but still make choices that may, or may not, be threatened by coercion. These include young children and some persons with severe mental health conditions, such as patients in a state of advanced dementia. Some non-human animals also arguably experience choice autonomy without agent autonomy, as when my dog decides to go for my carelessly placed lunch as opposed to some processed food from a tin (c.f. McMahan 1996; Kittay 2005).

The complete absence of choice autonomy is incompatible with agent autonomy. An agent subjected to coercion in every imaginable respect would be unable to genuinely exercise any capacity for self-governance. Yet some degree of agent autonomy is compatible with absence of choice autonomy in a wide range of respects, at least some of which are beyond ethical criticism. Thus, coercive interference in attempts to cause serious harm to others may restrict your choice autonomy but are not normally thought to

⁵ For an alternative account of autonomy that makes explicit use of earlier Aristotelian materials, see Radoilska [this volume].

undermine your capacity for self-governance. To the extent that it does, we are prone to accept it, or even question whether such actions should really be counted as autonomous at all (c.f. O'Neill 2002). To value agent autonomy does not mean to value it absolutely, or as a uniquely supreme value. Coercive interference with your choice of such various goods as hairstyle, hobbies, friends, career, or life-plan on the other hand, not only restrict your choice autonomy but are widely thought to undermine your capacity for self-governance. To the extent that it does, contemporary liberals are prone to condemn it. Failure to respect agent autonomy therefore stands in need of special justification. In the absence of such justification, any subversive interference with agent autonomy is expressive of an ethically undesirable form of disrespect towards the other as an independent locus of self-governance, authority and value.

It is natural to think that the notions of agent autonomy and choice autonomy stand in an asymmetrical justificatory relationship. One reason (perhaps the most important reason) one might think we should respect choice autonomy is that we should respect agent autonomy. This thought is clearly visible in contemporary justifications of advance directives for end-of-life medical treatment and in the controversial debate over how to respond to the wishes of severely demented persons (c.f. Dworkin 1994). In each case, the underlying thought is that whether or not we are required to respect choice autonomy depends on whether or not in doing so we will also respect agent autonomy, where agent autonomy is understood in terms of some prior exercise of genuine self-governance, such as publicly expressing one's reflective, or critical, endorsement of a life-plan. It is therefore unsurprising that much theoretical attention has been paid, first, to the

philosophical task of specifying the necessary and sufficient conditions for agent autonomy, and second, to the empirical task of establishing whether patients in some given medical condition (e.g. advanced dementia, severe personality disorder, etc.) in fact meet these conditions. If they do, there is a deep ethical justification in favour of respecting choice autonomy where this is at all practical. And if they do not, there is no such justification, in which case paternalistic intervention or other forms of coercion might in principle be justified, either by appeal to the patient's own interest or to the interests of others.

I do not wish to suggest that this is a worthless project. It is, however, a project that has the potential to create both ethical and conceptual confusion, and arguably continues to do so. I have two reasons for making this claim. The first is that some of the most controversial debates concerning our ethical relations to people with serious mental health conditions are premised, either implicitly or explicitly, on the assumption that these people do *not* meet all the standard conditions for agent autonomy. Thus, contemporary disagreements about the ethics of advanced dementia and severe personality disorder, for example, are not always focused on whether or not patients who suffer from these conditions are capable of genuine self-governance. On the contrary, it is often assumed that they are not; one question being how to then evaluate their choice autonomy, and whether we should think of choice autonomy as generating substantial constraints on the treatment of such patients in the absence of some appeal to a prior, future or potential capacity for, or exercise of, self-governance. No real progress is made on this question by giving an account of the conditions for agent autonomy and then

showing that the patients in question fail to meet them.

Second, the claim that the notions of agent autonomy and choice autonomy stand in an asymmetrical justificatory relationship does not entail that the value of respecting agent autonomy is the only justification there could be for why we should want to respect choice autonomy. There is a plurality of reasons why we might want to respect the choice autonomy of someone who is not agent autonomous, some of which essentially involve the idea of agent autonomy and some of which do not (c.f. Kittay 2005; Nussbaum 2009). Here I shall mention eight (the list is not exhaustive). First, you might want to respect the choice autonomy of someone because you believe that by doing so you will help them to become agent autonomous. Some people may adopt this attitude toward their children or people with curable mental health conditions. Second, you might want to respect the choice autonomy of someone because you believe that by doing so you will arrest or delay their progressive loss of agent autonomy. Some people may adopt this attitude toward people with advancing dementia. Third, you may want to respect the choice autonomy of someone because you believe that doing so is expressive of respect for the person they once were (and to a decreasing, but not negligible, extent still are). This is a potentially more problematic example, but any satisfactory account of our ethical relationship to ageing and death needs to have something to say about it. Fourth, you might want to respect the choice autonomy of someone because you believe that not doing so would be unpleasant or frustrating for them. This thought is arguably independent of ideas about agent autonomy. Yet it can be shown to play a central part in discussions of mental health where the values of agent autonomy and choice autonomy

are believed to conflict, including cases of patients with reduced mental capacities who express preferences that conflict with prior agent autonomous requests, such as an advanced directive.⁶ Fifth, you may want to respect the choice autonomy of someone because you believe that it is in their best interest to be given the freedom to choose, either in this case or in general. This is perhaps the most obvious and uncontroversial way in which respect for choice autonomy can be justified without appealing to agent autonomy. Sixth, you may want to respect the choice autonomy of someone because you believe that doing so is in the general interest, where this includes the interests of anyone else likely to be affected by the decision in question. Thus, it might be thought that drawing a sharp line around agent autonomy as a source of constraints against coercive interference with individual choice may have detrimental effects on respect for persons in general, leading to unacceptable forms of stigmatisation, repression or violations of basic legal rights.⁷ Seventh, you may want to respect the choice autonomy of someone because

⁶ The conditions for enforcing an advanced directive in the UK include the qualifications that no advance directive can be used to i) refuse basic nursing care essential to keep a person comfortable, such as washing, bathing and mouth care; ii) refuse the offer of food or drink by mouth; iii) refuse the use of measures solely designed to maintain comfort, such as painkillers, iv) demand treatment that a health care team considers inappropriate, v) ask for anything that is against the law, such as euthanasia or assisted suicide (Alzheimer's Society 2010).

⁷ In this context it is worth mentioning that at some times and places (including some places in the present) the standard treatment of people with serious mental health conditions has been worse than the treatment of many animals (c.f. Foucault 1989).

they manifest a number, but not all, of the standard features associated with agent autonomy. Some of these features will be intrinsic to the person so respected, such as having or expressing a recognisable outlook on their existence, however unstable or fragmentary. Other features will be extrinsic to the person so respected, such as being someone with whom one stands in an intimate, or otherwise special, relationship of familiarity or dependence. Eighth, you may want to respect someone's choice autonomy because he or she is another human being. Once more, this may come down to the possession of intrinsic characteristics by the person in question, such as human sentience or human form. Yet it may also be a question of extrinsic characteristics, including the relations in which they either stand or have stood to other human beings. Such forms of respect should not necessarily be thought of as respect for a second-rate form of agent autonomy. We may have deep reasons to value the relevant features of the agency of another human being that do not boil down to our valuing them as potential or second-rate manifestations of self-governance in the standard sense invoked in liberal moral and political philosophy. A human being whose mental capacities does not fully meet all the criteria of genuine self-governance need not be thought of as a second-rate person any more than a traffic warden need be thought of as a second-rate policeman or an EU citizen claiming residential rights in the UK need be thought of as a second rate Brit. The categories of agency we value as giving rise to ethically significant limitations on our treatment of others come in a cluster, not in a pair. They involve both intrinsic and extrinsic features of the agents in question. Their presence or otherwise are hardly, if ever, an all-or-nothing thing. They are also highly context sensitive. Thus, the virtues involved in caring for a friend or a family member are unlikely to be exactly the same as

the virtues involved in creating and operating public institutions designed to care for any patient meeting a certain diagnostic description. To think that the only feature of a human being that could justify respect for their choice autonomy is that of their agent autonomy is therefore both an overly simplistic and a potentially dangerous idea.

Agent autonomy and choice autonomy are two related, but separable, factors of ethical significance for our treatment of other human beings. One of the most important reasons for respecting choice autonomy is that in doing so we will often also be respecting agent autonomy. Yet respect for agent autonomy is not the only justification for respecting choice autonomy. It follows that we can have good reasons for respecting someone's choice autonomy even if they do not meet the conditions of genuine self-governance that are necessary for agent autonomy. This conclusion has potentially significant consequences for the ethics of mental health, including our attitudes towards people with mental disabilities and patients who suffer from advancing stages of dementia.

4. Autonomy as a value

Considerations of autonomy are sometimes thought to be in potential conflict with considerations relating to what is in someone's best interest. This conflict is sometimes phrased in terms of how autonomy relates to value or the promotion of desirable outcomes, whether the desirable outcomes in question are conceived of as broadly ethical or more narrowly prudential. In fact, there is a plurality of ways in which considerations of autonomy can be coherently related to considerations of value or the promotion of

desirable outcomes. Here I shall mention three.

First, autonomy is sometimes thought of as a conducive to the promotion of value or desirable outcomes (c.f. Keown 2002). There are at least two ways in which autonomy can be thought of as valuable in this way (the two are not mutually exclusive (c.f. Dworkin 1994)). First, it is sometimes argued that agents themselves are generally better placed to know what it is valuable for them to pursue, especially regarding what is in their own best interest. Preventing an agent's unconstrained pursuit of what they regard as being in their own best interest is therefore likely to be epistemologically counterproductive. Second, it is sometimes argued that agents themselves are generally better placed to successfully promote desirable outcomes, especially regarding what is in their own best interest. Preventing an agent's unconstrained pursuit of what they regard as being in their own best interest is therefore likely to be pragmatically counterproductive. In each case autonomy is valued as a means to the promotion of some desirable outcome, such as the agent's well-being. The issue of whether to respect someone's autonomy so valued then comes down to such questions as whether or not the agent in question is, in fact, better placed to know what is in his or her own best interest and/or whether or not the agent is, in fact, better placed to act in his or her own best interest. These claims are highly sensitive to context. Thus, it might be argued that with respect to some agents, such as patients suffering from severe mental disorders, paternalistic intervention is a more reliable mechanism for the promotion of the relevant desirable outcomes. If so, in these cases appeals to the value of autonomy do not gain support from appeals to the conduciveness of autonomy to the promotion of desirable

outcomes.

The claim that autonomy is conducive to the promotion of desirable outcomes can be understood as invoking both agent and choice autonomy. An agent's exercise of self-governance can be thought of as conducive to the promotion of desirable outcomes for at least three reasons. First, an agent's reflective endorsement of a practical option is often evidential of that option being valuable, such as when someone decides upon reflection that participating in a dangerous sport either is, or is not, for them. Second, reflective endorsement sometimes implies that the option so endorsed is valuable. Arguably, some things in life are good for you at least in part because you reflectively choose them. Third, being substantially self-governing entails a capacity for carrying one's reflectively endorsed plans out in practice, thereby enhancing the probability of realising the desirable outcomes in question. When autonomy is valued as conducive to value in any of these three ways, it is therefore often the idea of agent autonomy that is the focal point of valuation, choice autonomy being valued as a necessary condition for agent autonomy. The basic ability to act without coercive interference is at best contingently related to the promotion of desirable outcomes, being equally compatible with actions that frustrate the promotion of valuable states of affairs as with actions that realize them. It does not follow, however, that the only way for choice autonomy to be valuable as conducive to the promotion of desirable outcomes is as a necessary condition for agent autonomy. As I argued in the previous section, respect for choice autonomy can be valuable for reasons unrelated to agent autonomy, such as avoidance of pain and frustration that would reduce the agent's overall well-being. It might be tempting to think that autonomy being

conducive to the promotion of desirable outcomes is inevitably a matter of agent autonomy being so conducive if we focus exclusively on the epistemological aspect of autonomy being conducive to desirable outcomes. This temptation ought to disappear once we remember that something can be conducive to the promotion of desirable outcomes without being epistemologically so conducive.

Second, autonomy is sometimes thought of as itself valuable as an end, and not merely as a means to the promotion of desirable outcomes. Thus understood, it is natural to think that agent autonomy is valuable to a very high degree. On at least one influential conception of the self-governing individual as a primary holder of rights and duties, the failure to respect agent autonomy is the failure of treating someone with the basic respect they merit in virtue of being a locus of continuous and self-conscious rational agency (c.f. Dworkin 1994). Thus understood, respect for agent autonomy is likely to be assigned a fundamental place in any broadly liberal account of medical ethics. Yet choice autonomy can also be valued for its own sake. The fact of coercion is itself often a negative feature of action considered from the first-personal point of view. Thus, it is perfectly coherent to think that the basic ability to choose between available options is preferable to coercion, even for beings, such as some non-human animals, that have no capacity for agent autonomy. Yet the value of agent autonomy is often thought of as trumping the value of choice autonomy where the two conflict, as in the famous case of Ulysses and the Sirens (c.f. Elster 1984). Faced with lethal temptation, Ulysses is willing to trade his choice autonomy for a limited period in order to protect his continued agent autonomy in the long run. More difficult cases arise where a capacity for choice autonomy is present but a

capacity for agent autonomy is not. Such cases include the treatment of patients with severe mental incapacities, such as advanced dementia. In these cases, it is more controversial to what extent a prior exercise of agent autonomy (including the endorsement of practical options involving a future self who will not be agent autonomous) can reasonably be thought to trump a current exercise of choice autonomy (c.f. Shiffrin 2004). We can only hope to get clear about such cases if we consider the fact, mentioned in the previous section, that the features associated with agent autonomy come in clusters, have both first-personal and third-personal aspects, and are not always present or absent as an all-or-nothing thing (c.f. Heal [this volume]). Thus, you may want to respect the choice autonomy of someone because they manifest some, but not all, the features associated with agent autonomy. Furthermore, in doing so you may not be interested in these features exclusively, or even primarily, to the extent that they serve as markers for an ideal, but imperfectly realized, condition of autonomous agency. In doing so, your aim can be to respect someone else as exactly the kind of agent they actually are.

To assign agent autonomy or choice autonomy intrinsic value as ends opens up the question of how to compare such autonomy with other values, including personal well-being or impartial goods. Choice autonomy certainly, and agent autonomy probably, is not guaranteed to be the most important value in every situation where it conflicts with other values of these kinds. Furthermore, if the ethical significance of autonomy in either sense were exhausted by its role as a value understood thus far, it should be legitimate to think of this value in broadly consequentialist terms. On broadly consequentialist terms, the value of autonomy belongs on a scale of goods to be valued along with other ethically

significant goods. Thus, it must always be legitimate in principle to override autonomy in one case in order, for example, to respect more autonomy overall. Perhaps there is no principled obstacle to taking this consequentialist approach to choice autonomy. Yet for agent autonomy, considered as the primary locus of certain rights and duties, this consequentialist model arguably fails to do full justice to the idea that respect for an individual's capacity for self-governance ought to constrain our treatment of them in promotion of either their own or impartial goods (c.f. Scheffler 1988). This fact suggests that thinking of autonomy as either an instrumental or an intrinsic value in the sense discussed so far fails to capture the full ethical significance of considerations of autonomy, not only in the context of mental health, but also in the context of ethical and political philosophy more generally.

Third, autonomy is sometimes thought of it as a (deontological) constraint on the promotion of desirable outcomes (c.f. Nozick 1974). When thought of in this way, autonomy is not one further value to be weighed on a scale of goods, but instead an independent prohibition on, or obstacle to, taking certain means to promote otherwise desirable outcomes. As conceived by contemporary liberals, rights are paradigmatic candidates for the status of constraints in this sense. Thus, an individual's right to physical integrity is said to prohibit the taking of certain means to promote desirable outcomes, including the prevention of harm to that very individual. Hence the widely legislated duty to seek informed consent for invasive medical procedures. Yet constraints need not be absolute. Thus, an individual's right to physical integrity can be trumped by considerations of value in special circumstances, such as when the individual is a lethal

threat to himself or others, and/or is unable to control or otherwise exercise the relevant rights. Hence the widely legislated freedom to perform certain invasive procedures on patients whose lives are at risk and/or unable to consent. The presence of a constraint creates a special demand for ethical justification, yet not one that is obviously reducible to a matter of weighing values against each other on a scale of goods.⁸

Given the close association between the idea of a constraint and the idea of a right it is natural to think that the distinctive ethical significance of autonomy is partly to be explained in terms of the existence of constraints against the imposition of the will of some person(s) on that of another. In particular, it is natural to think that respect for agent autonomy imposes a substantial constraint that prohibits action to promote desirable outcomes where such action conflicts with the exercise genuine self-governance. To think of agent autonomy in this way is to understand the constraints it generates as an important part of what we mean by saying that there is a special kind of respect that we owe to persons as self-governing rational agents, a respect we arguably fail to show if we are always prepared to put autonomy on a scale of goods along with other values. It is much less plausible to think of choice autonomy as a substantial constraint in this sense, except where it functions as a necessary condition of respecting agent autonomy.

⁸ The presence of a constraint can be thought to involve a lexical ordering (where no amount of some particular value can ever trump a certain constraint) or the imposition of some significant threshold (where no amount of a particular value is taken to weigh against a constraint up to a certain point of seriousness). For further discussion, see Scheffler 1988.

Absence of coercion as such may be a genuinely desirable thing in itself, but considered in isolation from the content of the options at stake, the effects of its exercise on the interests of those affected, or its embedding in a psychological profile of a being that exhibits a set of features that we can see as meriting a distinctive form of respect, the appeal to choice autonomy itself is arguably too thin a basis on which to ground the kinds of constraints we normally associate with the rights and duties that are thought to characterize the more sophisticated ethical relationships we value as holding between paradigmatically moral persons (c.f. O'Neill 2002).

Given my remarks in the previous section, however, the significance of this point should not be exaggerated. Its connection with agent autonomy is not the only thing that could explain the imposition of constraints against the infringement of choice autonomy. Having their autonomy of choice frustrated can have seriously negative effects on someone's well-being and thereby frustrate interests that are sufficiently important to justify the imposition of duties on others to protect them. Thus understood, a constraint to respect choice autonomy would be associated with a species of individual rights that do not presuppose that the subject of those rights can also be a duty holder and thereby capable of genuine self-governance (c.f. Raz 1986). If so, there can be constraints against the infringement of choice autonomy even where agent autonomy is absent, such as in the case of patients with various kinds of mental illness and incapacity. Thus, it could be detrimental to the well-being of a severely demented patient to frustrate their preferences among a wide range of options, including some options between treatments or available medical procedures. According to the conception of choice autonomy just outlined, this

would then count against pursuing a course of action running contrary to those preferences, even where that course of action is favoured by a previous, and agent autonomous, advance directive. To what extent, if ever, considerations of choice autonomy should trump considerations of agent autonomy would then come down to a question about the conflict between two different kinds of right, namely those that derive from our nature as self-governing agents on the one hand, and those that derive from our nature as agents with distinctive and ethically significant perspective on the world on the other. Given the aforementioned asymmetry in justification between agent autonomy and choice autonomy it may be tempting to think that any rights derived from the former source will necessarily trump rights derived from the latter. As it happens, I think this view is too simple. I shall not, however, argue for this claim here (c.f. Dresser 1995; Shiffrin 2004).

Autonomy can be coherently thought of as an instrumental value, an intrinsic value, and also as a constraint on the promotion of value. On each model, agent autonomy is an uncontroversial source of ethical significance. Choice autonomy is ethically significant in the first two ways. And although its intrinsic ethical significance may not suffice to generate substantial constraints on promotion of value considered on its own, it is also possible to think that there can be genuine constraints on infringing choice autonomy, some of which are grounded independently of the fact that choice autonomy is a necessary condition for agent autonomy. We should therefore not think that choice autonomy is in principle too unsuited to justify genuine constraints on ethically permissible action wherever agent autonomy is absent, as in the case of some persons

who suffer from serious mental disorder or incapacity.

5. *Concluding remarks*

I have shown how considerations of autonomy normally play different roles in ethical thought from a first-person as opposed to an other-person perspective on practical reasoning. I have also shown that considerations of autonomy can be usefully distinguished into some that focus on a person's capacity for substantially self-governing agency and others that focus on an intellectually less demanding notion of voluntary choice or action. Finally, I have distinguished between three different ways of thinking about autonomy as a value, namely as a means to the promotion of desirable outcomes, as a desirable outcome in itself, and as a constraint on the promotion of desirable outcomes. I have argued that when these different aspects of claims about autonomy are kept distinct there is conceptual space for a view according to which we can reasonably consider ourselves to be under a duty to respect the autonomy of a person who does not have the capacities we normally associate with substantial, or genuine, self-governance.⁹

⁹ I am grateful to Lubomira Radoilska and the participants in her Cambridge seminars on autonomy and mental health, as well as to the audience of the January 2010 CRASSH conference on the same topic, for observations that have aided me during the writing of this paper. I am also grateful to Ben Colburn and Gemma Mitchell for many informative discussions in the context of their own work on the nature and value of autonomy.

References

Alzheimer's Society. 2010. Factsheet on Advance Directives.

<http://www.alzheimers.org.uk/factsheet/463>. Accessed 21 August 2010.

Arpaly, N. 2003. *Unprincipled Virtue*. Oxford: Oxford University Press.

Beauchamp, T. L. 2005. Who Deserves Autonomy, and Whose Autonomy Deserves Respect? In *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. J. Stacey Taylor (ed), 310-325. Cambridge: Cambridge University Press.

Christman, J. 1988. Constructing the Inner Citadel: Recent Work on Autonomy. *Ethics* 99: 109-24.

Darwall, S. 2006. *The Second Person Standpoint: Morality, Respect & Accountability*.

Cambridge MA: Harvard University Press.

Dresser, R. 1995. Dworkin on Dementia. *Hastings Center Report* 25: 32-38.

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

Dworkin, R. 1994. *Life's Dominion*. New York: Vintage.

Elster, J. 1984. *Ulysses and the Sirens*. Cambridge: Cambridge University Press.

Foucault, M. 1989. *Madness and Civilization*. London: Routledge.

Frankfurt, H. 1987. *The Importance of What We Care About*. Cambridge: Cambridge University Press.

Heal, J. [this volume]. Mental Disorder and the Value(s) of 'Autonomy'. In *Autonomy and Mental Health*, ed. L. Radoilska, [insert page numbers]. Oxford: Oxford University Press.

Holroyd, J. [this volume]. Clarifying Capacity: Value and Reasons. In *Autonomy and Mental Health*, ed. L. Radoilska, [insert page numbers]. Oxford: Oxford University Press.

Keown, J. 2002. *Euthanasia, Ethics and Public Policy*. Cambridge: Cambridge University Press.

Kittay, E. 2005. At The Margins of Moral Personhood. *Ethics* 116: 100-131.

Korsgaard, C. 2009. *Self-Constitution*. Oxford: Oxford University Press.

McMahan, J. 1996. Cognitive Disability, Misfortune and Justice. *Philosophy and Public Affairs* 25: 3-35.

Nozick, R. 1974. *Anarchy, State and Utopia*. Oxford: Basil Blackwell.

Nozick, R. 1997. *Socratic Puzzles*. Cambridge MA: Harvard University Press.

Nussbaum, M. 2009. The Capabilities of People with Cognitive Disabilities. *Metaphilosophy* 40: 331-351.

O'Neill, O. 2000. *Bounds of Justice*. Cambridge: Cambridge University Press.

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

Radoilska, L. [this volume]. Autonomy and the Freedoms that Matter. In *Autonomy and Mental Health*, ed. L. Radoilska, [insert page numbers]. Oxford: Oxford University Press.

Rawls, J. 1970. *A Theory of Justice*. Cambridge MA: Harvard University Press.

- Raz, J. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge MA: Harvard University Press.
- Scheffler, S. (ed). 1988. *Consequentialism and Its Critics*. Oxford: Oxford University Press.
- Scoccia, D. 1990. Paternalism and Respect for Autonomy. *Ethics* 100:318-334.
- Shiffrin, S.V. 2004. Autonomy, Beneficence, and the Permanently Demented. In *Dworkin and His Critics*. ed. J. Burley, 195-217. Oxford: Blackwell.
- Singer, P. 2009. Speciesism and Moral Status. *Metaphilosophy* 40: 567-581.
- Waldron, J. (ed). 1984. *Theories of Rights*. Oxford: Oxford University Press.
- Wikler, D. 1979. Paternalism and the Mildly Retarded. *Philosophy and Public Affairs* 8: 377-392.
- Williams, B. 2006. The Human Prejudice, in *Philosophy as a Humanistic Discipline*. A. W. Moore (ed), 135-152. Princeton: Princeton University Press.