Capacity, value neutrality and the ability to consider the future¹

Jillian Craigie University College London

Abstract

Calls for the adoption of a universal capacity approach to replace dedicated mental health law are motivated by the idea that the measures designed to protect patient autonomy in legislation such as the Mental Capacity Act 2005 should apply to everyone, including people with a psychiatric diagnosis. In this article it is argued that a diachronic perspective on questions of mental capacity is necessary if capacity law is to play this broader role, but that employing this perspective in assessments of capacity undermines central patient autonomy preserving features of the legislation, which presents a moral dilemma.

According to the Mental Capacity Act 2005 (MCA) in England and Wales, as well as similar law in other jurisdictions, treatment of a patient without their consent is justified only if their ability to make or express the relevant decision is compromised. This approach is designed to protect a patient's right to have their treatment decisions and motivating values respected, especially where the doctor disagrees about what is in the patient's best interests.² In this way, at least in theory, judgments of mental capacity are made largely independently of the treatment option that was chosen, and therefore of opinions about the choice. In adopting such an approach the law aspires to a kind of value neutrality.³

In addition to law based on the concept of mental capacity, many jurisdictions have a separate area of law that concerns detention and treatment in the context of psychiatric illness, which is very different in this respect. In England and Wales, the Mental Health Act 1983 (MHA) is explicitly value-laden, with a focus on the health and safety of the patient. Under this law a person can be treated for a mental disorder against their will, despite being judged to have the capacity to refuse treatment.

The moral tension that arises is that if respecting patient autonomy requires value neutrality when decision-making liberty is being curtailed, surely this must equally apply to psychiatric patients.⁴

Thanks go to Natalie Banner, Fabian Freyenhagen, Jonathan Glover, Gareth Owen, Wayne Martin, Genevra Richardson, Katie Steele, George Szmukler and Jo Wolff for their very helpful comments on earlier drafts of this article. Versions of the article were given at the Normative Neutrality and the Theory of Autonomy conference at the University of Essex, the Annual Conference of the Association for Legal and Social Philosophy at Queen's University Belfast, and the Medical Law and Ethics Research Seminar at King's College London, which helped a great deal its development. The article was written with the support of the Wellcome Trust [094910].

² When the decision is a request for treatment that the doctor believes is not in the patient's medical interests, rather than a refusal, this raises further issues. See, for example, Schwartz (1999).

³ Throughout this article I will be referring to value neutrality only in this narrow sense.

⁴ For an approach that emphasises patient autonomy as a guiding principle in the context of mental health law, see the *Report of the Expert Committee* (Department of Health, 1999); and in the context of the state of Victoria, Rees (2009).

This tension has been brought to the fore by the UN Convention on the Rights of Persons with Disabilities 2006, with which the MHA is seemingly not compliant due to its grounding of detention and treatment without consent in psychiatric diagnosis (Bartlett, 2012; Richardson, this volume). Concern about discrimination on the basis of psychiatric diagnosis has also been the central motivation behind calls for the fusion of mental health and capacity legislation, which would do away with dedicated mental health law (Szmukler and Dawson, 2011).

However, proposals to abandon mental health laws raise the question of whether a legal structure based on the concept of mental capacity, with its value-neutral approach, would sufficiently allow the state to intervene paternalistically in cases of psychiatric illness. Contexts in which patients could be judged to have mental capacity but are arguably not in a position to be making those decisions have been highlighted in anorexia nervosa (Tan, Hope, Stewart and Fitzpatrick, 2006; Hope, Tan, Stewart and McMillan, this issue; Carney, 2009) and mood disorders (Rudnick, 2002; Elliott, 1997). In the background is a long-standing debate about whether standard tools for assessing mental capacity adequately capture the problems of decision-making that seemingly justify the transfer of a decision to a surrogate (see, for example, Charland, 1998; Appelbaum, 1998; Breden and Vollmann, 2004; Okai et al., 2007).

This article explores whether conceiving of mental capacity in diachronic terms is consistent with the value-neutral aspirations of this area of law, and whether such an approach might enable mental capacity law to better deal with the kinds of hard cases in question. It is concluded that considerations derived from a diachronic perspective are relevant to questions of treatment decision-making liberty – especially in the context of mental disorders – but that they are difficult to reconcile with a value-neutral approach to identifying incapacity.

Mental capacity and the ability to consider the future

According to the MCA, only a compromised ability to make a decision in accordance with certain procedural norms, loosely defined with reference to the ability to 'understand', 'retain' and 'use or weigh' relevant information, or an inability to communicate a decision, properly grounds judgments of incapacity (s. 3(1)). In addition, the identified problems of decision-making must be attributed to 'an impairment of, or disturbance in the functioning of, mind or brain' (s. 2(1)).

This law aims to protect the individual's right to pursue their own ends by adopting a capacity test that focuses on the decision-making process rather than the substance of the patient's decision. Influential liberal thinking holds that it is not the place of the law to pursue certain ends – rather, the law should aim to provide a space for individuals to pursue their own ends (Plant, 2011) - and this position is reflected in English common law relating to mental capacity:

'[In assessments of capacity] it is most important that those considering the issue should not confuse the question of mental capacity with the nature of the decision made by the patient, however grave the consequences. The view of the patient may reflect a difference in values rather than an absence of competence and the assessment of capacity should be approached with this firmly in mind.'5

Whether the patient's decision is consistent with proposed substantive norms in relation to what they should pursue - for example that people should pursue good health, or that they should want to live - is said to be beyond the proper scope of the incapacity test. In adopting this position, English law allows that even the most controversial treatment decisions might be

NHS Trust v. T (Adult Patient: Refusal of Medical Treatment) [2004] EWHC 1279 (Fam).

explained by the fact that the patient's evaluation of the situation diverges from a clinical perspective, and perhaps many people's evaluation of the situation.

In practice then, questions of mental capacity might be thought of as boiling down to a question about whether a decision can be explained in terms of the patient's particular motivating commitments (their desires, values, projects, ideas about a good life); or whether the decision is properly explained in terms of a problem in decision-making, which is due to a dysfunction of mind or brain. Is the patient's controversial decision understandable in the light of their commitments? Or is the patient's ability to understand, retain, and use or weigh what the likely costs and benefits of treatment will be for them, currently compromised?

Scrutinizing a patient's commitments in the context of a capacity assessment raises moral concerns because the possibility that a choice can be explained by idiosyncratic commitments plays a significant role in protecting the patient's right to make controversial decisions. Perhaps the most straightforward approach to conceiving of commitments, and the one most likely to protect this right, is to say that the person's commitments are their current commitments – those they recognize at the time of the decision. The problem with conceiving of commitments in this way is that it fails to recognise certain problems of practical decision-making to do with motivational coherence over time. Someone with bipolar disorder, for example, might clearly be choosing or acting in accordance with their current desires during a manic phase, but this perspective fails to capture what seemingly goes wrong with practical decision-making in such cases. As a result, if this understanding of commitments is adopted in assessments of mental capacity the reach of this area of law arguably does not extend as far as it should - a person may be judged to have mental capacity where this seems doubtful. Psychopathologies characterised by this kind of problem seem better accounted for in a legal structure that conceives of mental capacity, and in particular, commitments, in diachronic terms.

A diachronic perspective might also help illuminate some other hard cases that are not normally characterised in terms of problems of motivational coherence over time. For example, someone with anorexia nervosa may be choosing rationally in accordance with their current commitments especially a strong desire not to gain weight, or to exert extreme control over what they eat - and so may be judged to have capacity on these grounds. However, an assessment of their evaluative capacities in diachronic terms might yield the conclusion that their decision-making ability is seriously compromised (Craigie, 2011). Their future self may come to the view that they should have been motivated in ways that they could not recognise or could not act on at the time, and this might be a fairly robust feature of people with a diagnosis of anorexia nervosa.⁶ This perspective in relation to questions of mental capacity chimes with the suggestion that particularly in psychiatry, 'colloquialisms [such as] "you will thank me later" sometimes become the unwritten rules within which the merits of a patient's consent are assessed' (Gostin, 1981, p. 742).7

For some evidence that this is the case, see Tan et al. (2006).

The Court of Protection judgment in the recent case of A Local Authority, Ev. A Health Authority – concerning a refusal of treatment in the context of anorexia nervosa - raises the issue of whether the patient would in retrospect be thankful for forced treatment. It is noted that E's parents, 'emphasise the long-term nature of the problem and the very poor quality of life that E has had in recent years. They ask: if E survives, having been forced to eat, will she thank us in five years' time?' (para. 36). In relation to the decision that that forced treatment is in E's best interests, the judge gives significant weight to the idea that 'We only live once - we are born once and we die once - and the difference between life and death is the biggest difference we know. E is a special person, whose life is of value. She does not see it that way now, but she may in future' (para. 137). Although these issues are not raised in the judge's consideration of whether E had the capacity to refuse treatment, they are clearly present in the background. The judge's position that E's life is valuable, though she currently doesn't recognise this, seems closely connected to the grounds he gives for the finding that E lacked capacity at the time of the judgment - that for E the 'need not to gain weight overpowers all other thoughts' (para. 49). In the best interest decision the judge clearly holds that E

Such a perspective may also offer a solution in cases where the patient does not have a psychiatric diagnosis. Mackenzie and Watts discuss a case of a woman who was judged to have the capacity to refuse life-saving treatment, and subsequently died, in a situation where she had recently been left by her husband, following the amputation of both her legs. They argue that the trauma of these events and her subsequent inability to imagine a life without her partner clouded the woman's judgment in relation to treatment:

'The tragedy is that in this case therapy or treatment might have corrected her assumptions regarding future quality of life, and a decision to accept the lifesaving treatment might have resulted.' (2011, p. 33; for a full description of the case, see Halpern, 2012)

Differences between these cases suggest that taking a diachronic perspective in assessments of mental capacity allows for decision-making to be compromised in at least two ways. In the bipolar disorder and anorexia nervosa cases a diachronic perspective is used to argue that a person's evaluative capacities may be compromised - the capacities that determine the desires or values that motivate a decision. In the emotional trauma case, which might equally be applied to cases of depression, a diachronic perspective is used to identify a problem in the person's ability to assess whether what they want or value is achievable. According to Jodi Halpern's original analysis of the emotional trauma case, the woman should have been judged to lack capacity on grounds that her beliefs about the future were 'unresponsive to evidence', rendering her unable to 'think through alternatives' (Halpern, 2012, pp. 108, 114).

One problem with adopting a diachronic perspective in assessments of mental capacity is that whether questions of procedural rationality are properly understood in diachronic terms remains a contested issue in the fields of economics, psychology and philosophy. The functional element of the MCA's incapacity test concerns problems in meeting certain procedural requirements problems in the ability to understand, retain, use or weigh relevant information. Sceptics deny that at least some norms derived from a diachronic perspective fall into this category. They hold that such requirements are substantive rather than procedural in nature, so I turn now to this theoretical concern.

Are we rationally required to value the future?

Much of the normative and descriptive work on a diachronic perspective in personal decisionmaking has taken place in the fields of economics and psychology, but these fields have tended to adopt contrasting positions on the normative question. Economists often treat a person's weighting of the future as a preference like any other, in a framework that assumes that preferences are not available to rational criticism. Thomas Nagel describes a view that captures this position, using the term 'prudence' to refer to taking a diachronic perspective in decisionmaking:

'Prudence cannot on this view be explained merely by the perception that something is in one's future interest; there must be a desire to further one's future interests if the perception is to have an effect ... There seems little doubt that most people have the desire that makes prudence

is having problems to do with evaluation, and this may help explain his assessment that E was unable to weigh 'the advantages and disadvantages of eating in any meaningful way' (para. 49). Making a direct connection between mental capacity and the consistency of commitments over time, the judge's consideration of whether E had capacity when signing two advance directives refusing treatment, a great deal of weight is placed on the consistency of E's intentions over time (paras. 54-70) [2012] EWHC 1639 COP.

possible [however, according to this view we are not] in any sense required to possess the desires in question: consequently we are not required to act on the specified considerations. If one lacks the relevant desire, there is nothing more to be said." (Nagel, 1970, p. 28)

Psychologists, on the other hand, often assume that people are rationally required to give significant weight to the future personal consequences of choices, the strongest version being that people ought to be temporally neutral in their decision-making. It is widely accepted within this camp that there are at least some good reasons to give less weight to future consequences, particularly in order to take into account an expected depreciation in their future value. This might be relevant where money is concerned and inflation rates are high. However, foreseeable changes in one's circumstances can also be relevant: £50 might be much more valuable to a final year law student than £100 will be for them in the following year, when they reasonably expect to be working for a corporate lawyer. On these grounds it might be rational for the student to choose £50 now, foregoing £100 in a year's time. Likewise, it is argued that uncertainly with respect to future consequences can provide a good reason to choose a certain outcome now, over a more valuable but uncertain outcome in the future (Baron, 1994, pp. 513-16; Frederick, 2006). I am using the term 'temporally neutral' to refer to decisions that give equal weight to future and present utility, excepting the above kinds of considerations. The pressing question is why people ought to be impartial with respect to all parts of their lives once these kinds of consideration are taken into account.

The arguments used to defend this position in the psychological literature tend to take the form that a temporally neutral orientation advances the goals of practical decision-making. Choosing an outcome merely because it will happen sooner is said to be irrational because if one option 'achieves your goals better, then you should choose that one, regardless of when you decide' (Baron, 1994, p. 514). Using the classic marshmallow experiments as an example, if a child would rather eat two marshmallows than one, then they should choose to wait and be given two marshmallows sometime later rather than one marshmallow now – because overall this achieves the child's goals better.8 This kind of argument considers the question of goal satisfaction from a position that is not situated at a particular point in time, and prescribes that people ought to take this perspective in their decision-making: 'they should be just as concerned about themselves a year from now as they are about themselves this minute' (p. 516). Individuals should give weight to the implications for their future self, so the argument goes, for their own overall good.

Philosophers have defended this kind of position in similar ways. Nagel argues that prudence is a requirement of practical reason, which is secured because the personal consequences of a man's action concern 'his future' (1970, p. 42; his italics). According to Nagel, this constraint on practical decision-making follows from the person's awareness that they persist over time. It reflects the 'individual's conception of himself as a temporally persistent being: his ability to identify with past and future stages of himself and to regard them as forming a single life. Failure to be susceptible to prudence entails radical dissociation from one's future, one's past, and from oneself as a whole, conceived as a temporally extended individual' (p. 58).9 It is a mistake in practical reasoning because it comes 'at the cost of dissociation from one's temporally extended self' (p. 69).

I borrow this example from the work of Walter Mischel and colleagues. For an overview, see Mischel (2004).

John Rawls writes that 'pure time preference is irrational' because 'it means [the individual] is not viewing all moments as equally parts of one life' (1971, p. 295). Similarly, Christine Korsgaard writes: 'So long as I occupy this body and this life, I am this rational agent, the same one ... [It] is misleading to ask whether my present self has a reason to be concerned with my future selves. This way of talking presupposes that the present self is necessarily interested in the quality of present experiences and needs a further reason to care for more than that. But insofar as I constitute myself as a living agent living a particular life, I will not in this way oppose my present self to future ones. And so I do have a personal reason ... to care for my future' (1989, pp. 126-27).

The conception of self as extended throughout a human lifetime is powerful in the context of questions of procedural rationality because it offers grounds for claiming that a person ought to give weight to considerations that they may not currently care about. It offers an answer to what I have characterised as the economists' view, that if a person has no interest in the future personal consequences of their choice there is nothing more to be said, rationally speaking. Applied to the context of treatment decision-making, this kind of perspective is powerful because it offers grounds for calling a patient's mental capacity into question, with reference to their ability to consider their future and value it appropriately.

Rejections of this kind of rational requirement are therefore often based on a rejection of the idea of the self as something that necessarily extends throughout a lifetime, and Derek Parfit's account of personal identity is often used to make this argument (Baron, 1994, p. 516; Frederick, 2006; Maclean, 2006). Parfit holds that the persistence of a person over time must be understood in terms of their 'psychological continuity' - the connectedness of psychological features such as memories, character, interests and preferences across a lifetime (Parfit, 1983; 1984). On this view, whether the future inhabitant of one's body will be the same person can be a matter of degree. Psychological features sometimes change so radically over a human lifetime that a person now will not necessarily be the same person who occupies that body in the future. And not giving weight to the interests of a future self is irrational only to the degree that one shares the relevant psychological characteristics with that self.

More recently, Galen Strawson has defended a similar position in his rejection of narrative accounts of the self. Strawson argues that although experiencing oneself as one person throughout a lifetime - what he calls a 'Diachronic' self-experience - may be more common, some people experience themselves in what he describes as a more 'Episodic' way. Unlike the more Diachronic person, the more Episodic person has 'little or no sense that the self that one is was there in the (further) past and will be there in the future, although one is perfectly aware that one has longterm continuity considered as a whole human being' (Strawson, 2005, p. 65). '[Predominantly] Episodic individuals may sometimes connect to charged events in their pasts in such a way that they feel those events happened to them - embarrassing memories are a good example - and anticipate events in their futures in such a way that they think those events are going to happen to them - thoughts of death can be a good example" (p. 65). However, the Episodic person has no 'great or special interest in [their] past', and nor do they have 'a great deal of concern for [their] future' (p. 67). Strawson holds that even a strongly Episodic self is within the normal range of human experience, and that it is normatively on a par with a more Diachronic self-experience. On his view these are just different ways of experiencing and living a human life. 10

In summary, theorists such as Nagel assert that selves extend over a lifetime, and diachronic norms of procedural rationality are derived from this claim. Parfit and Strawson dispute this assumption, and conclude that disregard for the future personal consequences of a choice may simply reflect an alternative way of living a human life, at least when we consider the question of reasons for action in non-moral terms.^{II} However, all of the theorists canvassed thus far reason from premises about the extension of self over a lifetime, to conclusions about what reasons individuals have, and it will be argued here that this conception of the relationship between the requirements of practical rationality and the metaphysics of self is the wrong way around. As described above, the kind of requirement in question is often defended - by Nagel and others - on

¹⁰ There is at least some empirical support for his descriptive claim. One study found that people in the US experience themselves as least diachronic at beginning of adulthood (teens, 20s) and in later in life (50s, 60s), and most diachronic in their 30s and 40s (Frederick, 2003).

¹¹ This conclusion is explicitly endorsed by Parfit. I have also attributed this position to Strawson as it would seem to follow from his claims about an Episodic self-experience.

the grounds that not giving weight to the future undermines one's own interests. However, at times a different kind of argument is suggested in Nagel's work. The suggestion seems to be that a unified self over a lifetime is not something that is automatically present, but rather something that is generated by giving weight to future personal consequences in decision-making. The proposed requirement is justified, not because it deploys the correct conception of self, or because one's future self will otherwise regret the decision, but because adherence plays a role in bringing the self into being, by enabling the decision-maker 'to reach towards something outside himself (Nagel, 1970, p. 43).

Michael Bratman is particularly clear about the role that he proposes giving weight to the future plays in the generation of the self, and how this in turn grounds a diachronic perspective on questions of procedural rationality, so it is his view that I will focus on here. Central to Bratman's view is the idea that planfulness and what he calls 'self-governing policies' - which are integral to temporally extended agency - lie at the core of human agency because it is through these that a person's identity emerges. These processes are essential for understanding what it is for an agent to 'take a stand' - what it is for an agent to recognise certain desires as one's own - and it is this feature in particular that distinguishes human agency from the agency of other 'purposive' creatures (Bratman, 2000, pp. 50, 45).

Adopting broadly the same approach to personal identity as Parfit begins with, Bratman suggests that the forward-looking psychological ties that play a role in constituting a unified self over time can be generated through reflective and planning processes: I can help ensure appropriate psychological continuities and connections by sticking with and executing my prior plans and policies, and by monitoring and regulating my motivational structures in favor, say, of my continued commitment to philosophy" (2000, p. 45). Such processes 'play an important role in the constitution and support of continuities and connections characteristic of the identity of the agent over time'. 'Indeed,' Bratman suggests, 'this is what plans and policies are for' (2000, p. 47; his italics). By emphasising the role that planning and policy-making play in generating the psychological ties that are proposed to underpin personal identity over time, Bratman describes a picture of human agency in which the agent plays an active role. The temporally extended self develops through the agent's exercise of certain psychological capacities. On this point Bratman makes a significant departure from Parfit and Strawson, who seem to hold that the degree to which the relevant psychological ties extend over time is a fixed fact about the individual, from which rational requirements on their action may be derived.

The claim that engaging in planning and policy-making processes is rationally required is defended primarily on the non-instrumental grounds that these processes play a central role in developing and sustaining the self.12 Bratman's argument takes the form of an answer to the problem that agents who reflect on their first-order desires and form second-order desires do not yet exhibit what is distinctive about human agency, because there is nothing to distinguish these new desires as those that the agent identifies with; second-order desires seem to be 'just one more wiggle in the psychic stew' (Bratman, 2000, p. 38).¹³ The solution that Bratman offers is that where an agent stands with respect to a particular first-order desire is established through the exercise of planning and policy-making practices: 'Self-governing policies might, so to speak, crystallize pressures from various elements of one's psychic stew into a more decisive attitude that can, in

¹² Bratman also holds that there are instrumental reasons for engaging in such processes – reasons to do with the kinds of things one can do as a diachronic agent. Christine Korsgaard expresses a version of this form of response to Parfit, when she writes: 'In order to carry out a rational plan of life, you need to be one continuing person. You normally think you lead one continuing life because you are one person, but according to this argument the truth is the reverse. You are one continuing person because you have one life to lead' (1989, p. 113).

¹³ Here Bratman is giving a version of a standard objection to Harry Frankfurt's account of the will.

the relevant context, establish where one stands' (2000, p. 51). These mental processes that underpin temporally extended agency give authority to particular desires, so it is through these that the self – the distinctive feature of human agency – emerges. On this view, a compromised ability to engage in planning and policy-making processes threatens not the interests of the future self, but the very constitution of the self. This is what grounds the procedural requirement that one must give weight to the future personal consequences of actions. And this, we might suggest, is what makes the ability to consider and value the future an appropriate consideration to include in a mental capacity test.

Bratman's account offers a story about the self that fits well with a developmental understanding of personhood that is often appealed to in contemporary medical law and ethics. His account of the relationship between temporally extended agency and the self also fits well with the difficulties associated with certain psychopathologies. It is no accident, I suggest, that disorders such as dementia, bipolar disorder and depression are associated with seriously compromised capacities for diachronic agency. Among many other factors, in a complex picture, problems of diachronic agency surely make a contribution to our sense that these constellations of behavioural and experiential features are appropriate targets for the label of disorder. A sense of the self as compromised in the context of problems of diachronic agency is also found in first-person reports of dementia (Addis and Tippett, 2004; however, see Caddell and Clare, 2012) and depression (Fuchs, 2008; Ratcliffe, 2012). It would seem that the functioning of capacities for diachronic agency is connected to our sense of others as having well-functioning human agency, as well as our subjective sense of selfhood, and for these reasons I find it implausible that a strongly episodic self-experience is just a less common way of living a human life. Strawson defends a view along these lines, claiming that he himself is strongly Episodic (in his sense). Yet he has a successful academic career which could not have been achieved without meeting a wide range of commitments that each extended over considerable periods of time, for example in completing academic qualifications and writing books. It seems to me that these achievements evidence considerable skills in diachronic agency. While for Strawson a Diachronic self-experience is connected to a narrative conception of the self, diachronicity in the sense concerned here has no necessary connection to narrativity. A life lived in the moment could still be a life lived with diachronic agency, as I will argue at the end of the following section. I now turn back to questions of mental capacity and what implications a diachronic perspective such as Bratman's would have.

Including the ability to consider the future in assessments of mental capacity

The motivation for investigating these theoretical questions about procedural rationality was the idea that a diachronic perspective might enable a mental capacity test such as that adopted in the MCA to deal with certain hard cases, particularly in the context of mental disorder, without undermining its value-neutral approach. The arguments above suggest that a diachronic perspective on questions of mental capacity is consistent with a procedural understanding of the functional element of the MCA's incapacity test. Given this, could a diachronic understanding of what it is to have mental capacity help by justifying the inclusion of a requirement such as the ability to consider the future – as a part of what it means to be able to understand, retain, use or weigh information - in assessments of mental capacity?

Something like this requirement already seems to be assumed in the condition that the patient must understand the consequences of the treatment options. The relevant information necessarily concerns the future as well as the present, due to the fact that understanding 'the likely effects of deciding one way or the other' is a part of this requirement (Mental Capacity Act 2005 Code of Practice, s. 4.16). For example, a person with a severe learning disability might be judged to lack the capacity to consent to a blood test on grounds that they do not understand that the test is

necessary for their future physical well-being. The same might be said in a case of a delirious patient who is resisting treatment for a serious injury.

In such cases the person's connection to the world, in the sense of their ability to understand how things are likely to be in the future, seems compromised. Their ability to assess what to do is limited in a way that can be compared to John Stuart Mill's paradigm case of justified paternalistic intervention, where a person is about to cross a bridge that unbeknownst to them is unsafe. The person in the example does not understand that they risk falling into water if they attempt to cross. The widely accepted conclusion is that the person is not in a position to assess whether to cross the bridge, and this justifies the use of proportional physical force to stop them from crossing, when this is necessary to inform them about the state of the bridge.¹⁴

The same reasoning might be applied in the case of the woman who has recently suffered a serious emotional trauma, on grounds that her ability to imagine a range of future possibilities is severely limited (Halpern, 2012). The patient cannot currently conceive of alternative futures to the desperately unhappy one she imagines, and perhaps this justifies taking the decision out of her hands until she is seeing things more clearly. The problem that returns here is that understanding concerns not simply how things are likely to be, as in the fact that an attempt to cross the bridge is likely to result in a fall into water, or that a certain proportion of people who have both legs amputated learn to walk again using prosthetic limbs. It concerns understanding how the relevant state of affairs will be for this particular patient. This question about the future necessarily involves an evaluative judgment that, according to a value-neutral approach, only the patient is in a position to make.

So while it might sometimes be clear that a patient's ability to consider the future is compromised, for example when someone denies that their condition is life-threatening when it clearly is, in many cases assessing understanding in relation to the future will have no equivalent point of reference. In such a context there seems to be no standard against which the functional element of the mental capacity test – the part concerning the patient's ability to understand, retain, use or weigh information - can be directly assessed. Therefore, any conclusions drawn about this capacity would seem to be based on the fact that the patient has suffered an emotional trauma. Removing a patient's right to make their own treatment decision under these circumstances may seem justified given what we know about how people who lose limbs adapt, as reflected in self-reported measures of well-being, and given the known psychological effects of emotional trauma. However, an appeal to this kind of knowledge in an assessment of mental capacity would involve using generalised observations to draw a conclusion about a particular person's capacity to assess how things will be for them in the future. The legislation is clear that an assessment of decision-making ability should not be based solely on features such as a patient's diagnosis.15 This measure seems designed to preserve patient autonomy - it must be shown that this particular person's mental capacities are relevantly impaired, not merely that they belong to a diagnostic category. However, this requirement cannot be met in the case of this capacity, because

¹⁴ Mill, 1910/1972, pp. 151-52: 'for liberty consists in doing what one desires, and he does not desire to fall into the water.' Mill goes on in the following sentences to extend the example to a case of delirium. The analogy is closer in the case of a delirious patient than a person with a learning disability, assuming that the delirious patient will recover full consciousness and will want treatment for their injury. The case of the person with a learning disability is disanalogous in the sense that their incapacity for understanding is most likely permanent. In both cases, however, under English law the treatment decision would be transferred to a surrogate and made in the patient's best interests, taking into account their past and present wishes, feelings, beliefs and values. The assumption in both contexts is that the patient is not knowingly making a life-threatening decision - they do not wish to die - and it is this same kind of assumption that justifies intervention in the broken bridge example.

¹⁵ MCA s. 2(3)(b): 'A lack of capacity cannot be established merely by reference to ... a condition of his, or an aspect of his behaviour, which might lead others to make unjustified assumptions about his capacity.'

there is no way to directly assess it while remaining value neutral in accordance with the aspirations of this area of law.16

What about the possibility raised at the beginning of this article, that a person's evaluative capacities might be called into question when viewed from a diachronic perspective? Drawing on Bratman's account, planfulenss and self-governing policies are rationally required because of the role they play in developing and sustaining the self. Caring about the interests of the future inhabitant of one's body, and having at least some grip on what those interests will be, seem like prerequisites for engaging successfully in these processes. And perhaps this offers part of an explanation about how decision-making can be compromised in the context of disorders such as depression and bipolar disorder. Very crudely, the idea would be that as part of a much bigger picture the capacities necessary for diachronic agency are not functioning properly and this undermines the constitution of the self. I have argued that, understood in this way, these considerations are procedural rather than substantive in nature. But would the implementation of these considerations in practice nonetheless come into conflict with the goal of preserving patient autonomy?

Assessing a person's grip on what interests the future inhabitant of their body will have runs into the same difficulties that occurred in the context of assessing understanding in the emotional trauma case. Without making substantive assumptions there is no standard against which this capacity can be directly tested. Judging that a person is likely to lack these capacities in the context of a manic episode may seem justified, because of what we know about the trajectory of bipolar disorder that fairly reliably people's commitments during a manic episode radically undermine their overall interests. But whether this kind of consideration provides grounds for a judgment of incapacity in English law depends on what evidence is considered appropriate for assessing the functional element of the incapacity test. The MCA prohibits basing judgments of mental capacity merely on grounds of diagnosis. However, as Genevra Richardson points out, 'existing case law indicates that courts will take a flexible approach to the legal definition [of capacity] to enable them to reach their preferred outcome (Richardson, this volume, p. oo).¹⁷ This may mean that in practice significant weight is given to a person's diagnosis. However, this solution undermines the protections of patient autonomy conferred by making an individual assessment of functional capacities, rather than diagnosis, the proper grounds for infringements on self-determination.¹⁸

¹⁶ A similar argument seems to underpin ruling out the bizarreness of a decision as an appropriate basis for a judgment of incapacity. An appeal to the bizarreness of a decision might be justified in evidential terms, on grounds that on the whole people tend to value a narrow range of things within the full range of things they could value. If this is true, then the most likely explanation for any controversial decision is a problem in the patient's ability to understand or process information, rather than them having an unusual evaluation of the situation. However, the aim of the law to preserve the individual's right to pursue their own commitments would seem particularly relevant for those whose commitments are marginalised, either in relation to the medical establishment or to mainstream culture. So the claim that only a small minority of people who make controversial decisions will be doing so because of their unusual evaluation of the situation could be interpreted as merely highlighting the need to protect this group. It is because they are a minority group that their right to self-determination needs protecting, and this is what makes it inappropriate to base an incapacity judgment on the bizarreness of a decision despite this being evidentially justified.

¹⁷ See also an analysis of English mental capacity law by Jules Holroyd (2012, p. 159), which finds that 'meeting the conditions for capacity seems to require not merely that the agent hold some evaluative commitments . . . but also that the agent hold some specific and substantive evaluative commitments.'

¹⁸ The conundrum is poignantly described by the parents of E, a woman with anorexia nervosa, in the case A Local Authority, Ev. A Health Authority. 'It seems strange to us that the only people who don't seem to have the right to die when there is no further appropriate treatment available are those with an eating disorder. This is based on the assumption that they can never have capacity around any issues connected to food. There is a logic to this, but not from the perspective of the sufferer who is not extended the same rights as any other person' (para. 52).

A related possibility considered at the beginning of this article was that a diachronic perspective might be used to justify giving weight to a reasonable prediction of what the patient's future self would want in retrospect, in assessments of mental capacity. However, it does not follow from the view offered here, even given the sure knowledge that a patient's future self would retrospectively endorse a medical intervention, that their present self is irrational in refusing it. In the imagined case there is an unambiguous conflict between what the patient currently wants (no treatment) and what the future self will want in retrospect (treatment).¹⁹ Using this conflict to doubt the patient's current capacity to make the treatment decision assumes that the future self's perspective should be given greater weight in the current self's decision-making. However, there is nothing in the present view that would justify resolving the conflict in this way.²⁰ The proposed account requires that the agent exercise planfulness and self-governing policies in decision-making, which would seem to involve giving equal weight at most, not greater weight, to one's expected future desires. Moreover, humans have a well-established time preference for negative experiences to be in the past rather than the present or future, and this makes it easier for the patient to endorse involuntary treatment when it is in the past (Doherty, 2011). This would seem to weigh in favour of privileging the perspective of the self who is facing the infringement on liberty. In the absence of any reasons to believe that the future self's perspective should be privileged, this form of justification for doubting mental capacity looks to be question-begging. Its favouring of the future self's perspective already assumes that the current self's decision-making capacity is relevantly impaired. Favouring the future self's perspective might seem justified on grounds that, for example, evaluative capacities are often compromised in the context of certain diagnoses or during experiences such as emotional trauma. However, at least in theory, this evidence alone does not justify a judgment of incapacity under current English law, for reasons of preserving patient autonomy as argued above.

Finally, I turn to the question of caring about future personal interests, especially in the context of life-threatening choices. A central difficulty associated with including a consideration of this kind in questions of mental capacity is the problem of how to assess caring simpliciter, independently of caring about a particular outcome. According to the account offered here, procedural rationality requires that we care about the future personal consequences of choices, but what in the future we should give weight to is left open. Specifying this in assessments of mental capacity would mean that the test was not value neutral. The case of the committed smoker (exemplified by David Hockney²¹) can be used to illustrate the point. One criticism often levelled at smokers is that they are irrational because they do not care sufficiently about the future personal consequences of their smoking. However, this criticism assumes that future health, or a long life, is what the person should care about. Considering a person's decision to keep smoking from a value-neutral perspective, which

¹⁹ I will set aside difficult and important concerns about how we can know a person's future psychological states, and how the course of action taken in the present may influence their future assessment of the situation. For example, perhaps a patient will be more likely to retrospectively endorse the involuntary treatment if they are treated involuntarily; or vice versa. It looks like the future point of reference cannot be fixed, independent of the decision that is made about the patient's mental capacity.

²⁰ There has been extensive discussion of this issue in the medical ethics and law literature, particularly in the context of Ulysses contracts in psychiatry. For example, see Dresser (1984), Andreou (2008).

²¹ In response to campaigns against smoking in the UK, the artist David Hockney has passionately defended smoking as something that can compete with and, for him, outweighs the value of considerations of future health and a long life: 'Some people [like me] want to live and they don't want to live like you do. It doesn't matter if I die early.' 'Death awaits you whether you smoke or not' (BBC News, 2005). 'If [smoking] kills me next week, I don't care. Not everyone can live to be 100. My mother did and she told me it wasn't too much fun.' I am glad of the tobacco manufacturers. I am a big customer of theirs. They make a good vegetarian product' (Woolf, 2005).

makes no assumptions of this kind, is more difficult task. Are they being short-sighted? Or do they care about things other than health and long life in a way that makes their decision rational?

Bratman's account shows how asking this second question can be consistent with a diachronic perspective on rationality. It is clear that the committed smoker is very likely undermining the health of the future inhabitant of his body. However, it does not follow that he is not engaging in the psychological processes that underpin temporally extended agency. It may be a central feature of the person's life plan that he should not be overly concerned about living a long life. His deep and enduring commitment might be to a hedonistic lifestyle that is incompatible with being concerned about health in old age. For such a person, not living fully enough is a more likely source of future regrets than health-endangering behaviours. Caring about one's future might even require, in such a case, life-threatening or health-threatening choices. Parallel arguments could be made in the case of a decision by a Jehovah's Witness to refuse a blood transfusion; for a political activist's hunger strike; for a person who risks their life in an attempt to save a loved one; and a freediver's commitment to their dangerous extreme-sports lifestyle. These kinds of decision do not seem characterized by a lack of planfulness about the future. Rather than undermining the sense of oneself as 'a temporally persisting agent', decision-making in line with these kinds of commitment looks likely to contribute to the 'overlapping webs of cross-temporal connections and continuities' that are proposed to play a role in constituting human agency (Bratman 2000, pp. 46, 47). The future self taken into consideration in these cases may be a nearer self than in the case of a person who is invested in living a long life - because the artist who is a committed smoker, for example, might not expect to or want to live a long life – but that is part of their plan. Nothing in the above account enjoins us to aim for a life of a particular length. What it requires is that, at least to some degree, choices be made with a reflectivity about the future, such that insofar as we are agents that agency is temporally extended.²²

This perspective allows us to reconcile the ideas that mental capacity requires that we care about the future personal consequences of our choices, and that life-ending decisions can be made with mental capacity. However, against this theoretical background, resolving questions about whether a patient's capacity to care about the future is compromised would seem to require looking beyond their procedural capacities. Answers to this question will rest on views about whether or not the life-threatening decision is consistent with the person's commitments, understood from a diachronic perspective. These factors are taken into account in an assessment of best interests once it is decided that a patient lacks mental capacity (MCA s. 4(6)). However, if the ability to care about the future is included as part of what it means to have mental capacity, then consideration of a patient's commitments from a diachronic perspective will also be central to this prior decision.

²² It might still be objected that life-threatening choices jeopardise the conditions necessary for temporally extended agency, because life is necessary for agency of any kind; and that paternalistic intervention is justified when a patient's choice threatens the conditions that are necessary for future human agency. A similar argument is sometimes given to justify paternalism in cases where an individual makes a choice that will undermine their ability to be autonomous in the future – for example in the case of choices that will result in a person killing or crippling themselves. It is the value of personal freedom, normally used to justify an anti-paternalistic stance, which is given as a reason to intervene in cases where the conditions for future personal freedom are under threat. For example, see Korsgaard (1989). However, this would not seem to be an argument for the irrationality of such a choice on the view offered here. According to Bratman's account, rationality requires that weight be given to the future personal consequences of choices because it is through planning and policy-making processes (among other things) that personal identity emerges. However, it was argued above that this requirement is consistent with some life-threatening decisions. So to conclude that such decisions are nonetheless problematic on grounds of their threat to the conditions of human agency would seem to assume that continued human agency has a value that trumps the agent's commitments. This just looks to be a substantive claim based on the categorical value of human agency - that a person should want to live no matter what their conception of a good life is.

A dilemma about what respect for patient autonomy requires

It has been argued here that a diachronic perspective on mental capacity is relevant to questions about when an infringement on treatment decision-making liberty is justified. However, including considerations derived from a diachronic perspective in assessments of capacity comes into conflict with central patient autonomy-preserving features of the MCA. It's worth reflecting, then, on why a diachronic perspective might be morally relevant in this context. I will suggest that the answer lies in a common view about what makes patient autonomy valuable - what makes it the case that for the most part, people's treatment decisions, especially their refusals, ought to be respected.

The importance of allowing patients to be self-determining in relation to their own medical care is often justified along lines that it 'protects people's general capacity to lead their lives out of a distinctive sense of their own character, a sense of what is important to and for them' (Dworkin, 1994, p. 224). Different theorists argue for the value of patient autonomy on importantly different grounds, but a common theme is that treatment decisions should be respected because of their assumed connection to the self. However, in a case where human agency is seriously compromised in the way described by Bratman, selfhood is undermined. Capacities for diachronic agency that play a central role in establishing what's important for the person – where the person stands – are compromised. And in the context of such problems, the importance placed on patient autonomy in liberal societies would seem to give medical professionals a reason to do what they can to restore selfhood, perhaps even if this requires an infringement on personal liberty.²³ This might involve overriding a refusal of treatment, for example, though the justification for doing so would only warrant interventions of a particular kind – those that are likely to restore the general capacity that the principle of patient autonomy aims to protect.

This suggests that in some cases, especially in the context of mental illness, the importance placed on patient autonomy in contemporary medical ethics presents the law with a dilemma. On the one hand, it seems right that laws which license infringements on decision-making liberty should equally apply to everyone, not just to particular groups; and that if a patient has mental capacity in relation to a particular treatment decision then they should be free to direct the course of their medical care. Consequently, it seems we ought to do away with dedicated mental health laws that legalise detention and involuntary treatment without reference to mental capacity only in the context of a mental disorder. On the other hand, the way in which diachronic agency is compromised in some instances – especially in the context of mental disorder – seems to justify or perhaps even require an infringement on liberty to restore the self.24 These grounds are difficult to take into consideration in assessments of mental capacity without compromising features of the MCA that are designed to preserve patient autonomy. So while it might be possible to implement the MCA in a way that sufficiently accounts for the problems of decision-making and selfhood that are sometimes present in mental disorders, doing so comes at the price of what is seen to be a central virtue of the legislation.

Conclusion

In this article my focus was requirements of rationality derived from a diachronic perspective, and the question of what kinds of consideration this perspective would enable mental capacity assessments to include. It was an attempt to explore the limits of the MCA's incapacity test, given its value-neutral

²³ Along similar lines, others have argued that restoring autonomy that is being undermined by mental illness can constitute 'a reason for [coercive] medical treatment' (Sjöstrand and Helgesson, 2008, p. 117).

²⁴ See Edwards (2010) for a different argument to the same conclusion.

approach. It was argued that considerations derived from a diachronic perspective are relevant to questions of treatment decision-making liberty, and are consistent with a procedural understanding of the functional element of the test. However, in practice, appeal to these considerations in assessments of mental capacity undermines the role played by the functional element of the test in preserving patient autonomy. While involuntary treatment without reference to mental capacity is often assumed to undermine the principle of respecting patient autonomy, it was argued here that this is not necessarily so. When human agency is seriously compromised, preserving patient autonomy may sometimes require overriding a treatment decision to restore selfhood, on grounds that are not easily accounted for under the MCA without compromising the value-neutral aspirations of this area of law.

In considering a move to abandon dedicated mental health law, legislators must therefore resolve a bind. The central motivation for such a move is concern about unjustified infringements on liberty in the context of psychiatric diagnoses. The worry is that the protections of patient autonomy in the MCA do not currently extend to the treatment of mental disorders, because mental health legislation allows people to be treated for a mental disorder even if they refuse with mental capacity. However, considerations derived from a diachronic perspective that seem relevant to when an infringement on decision-making liberty is justified are especially present in mental disorder. The inclusion of these considerations seems necessary for adequately caring for those with a mental disorder under universal capacity legislation. Yet including these factors in assessments of mental capacity undermines the protections of patient autonomy that supporters of a shift to universal capacity legislation hope to gain.

References

- ADDIS, Donna Rose and TIPPETT, Lynette J. (2004) 'Memory of Myself: Autobiographical Memory and Identity in Alzheimer's Disease', Memory 12(1): 56-74.
- ANDREOU, Chrisoula (2008) 'Making a Clean Break: Addiction and Ulysses Contracts', Bioethics 22(1): 25-31.
- APPELBAUM, Paul (1998) 'Ought We to Require Emotional Capacity as Part of Decisional Competence?', Kennedy Institute of Ethics Journal 8(4): 377-89.
- BARON, Jonathan (1994) Thinking and Deciding. Cambridge: Cambridge University Press.
- BARTLETT, Peter (2012) 'The United Nations Convention on the Rights of Persons with Disabilities and Mental Health Law', Modern Law Review 75(5): 752-78.
- BBC NEWS (2005) 'Hockney Leads Smoking Ban Protest.' Available at: http://news.bbc.co.uk/1/hi/ uk politics/4288918.stm.
- BRATMAN, Michael (2000) 'Reflection, Planning, and Temporally Extended Agency', Philosophical Review 109(1): 35-61.
- BREDEN, Torsten Marcus and VOLLMANN, Jochen (2004) 'The Cognitive Based Approach of Capacity Assessment in Psychiatry: A Philosophical Critique of the MacCAT-T', Health Care Analysis 12 (4): 273-83.
- CADDELL, Lisa S. and CLARE, Linda (2012) 'Studying the Self in People with Dementia: How Might We Proceed?' Dementia, advance online publication, doi: 10.1177/1471301211418486.
- CARNEY, Terry (2009) 'Anorexia: A Role for Law in Therapy?' Psychiatry, Psychology and Law 16(1): 41-59.
- CHARLAND, Louis C. (1998) 'Appreciation and Emotion: Theoretical Reflections on the MacArthur Treatment Competence Study', *Kennedy Institute of Ethics Journal* 8(4): 359–76.
- CRAIGIE, Jillian (2011) 'Competence, Practical Rationality and What a Patient Values', Bioethics 25(6): 326-33.

- DEPARTMENT OF HEALTH (1999) Review of the Mental Health Act 1983 Report of the Expert Committee. London: Department of Health.
- DOHERTY, Tom (2011) 'On Whether to Prefer Pain to Pass', Ethics 121(3): 521-37.
- DRESSER, Rebecca (1984) 'Bound to Treatment: The Ulysses Contract', Hastings Center Report 14(3):
- DWORKIN, Ronald (1994) Life's Dominion. New York: Vintage Books.
- EDWARDS, Craig (2010) 'Beyond Mental Competence', Journal of Applied Philosophy 27(3): 273-89.
- ELLIOTT, Carl (1997) 'Caring About Risks', Archives of General Psychiatry 54: 113–16.
- FREDERICK, Shane (2003) 'Time Preference and Personal Identity', in George Loewenstein, Daniel Read and Roy Baumeister (eds), Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice. New York: Russell Sage Foundation, 89–113.
- FREDERICK, Shane (2006) 'Valuing Future Life and Future Lives: A Framework for Understanding Discounting', Journal of Economic Psychology 27: 667–80.
- FUCHS, Thomas (2008) 'The Phenomenology of Body and Space in Depression', Philosophy 36(1): 69-84. GOSTIN, Larry O. (1981) 'Observations on Consent to Treatment and Review of Clinical Judgment in Psychiatry: A Discussion Paper', Journal of the Royal Society of Medicine 74: 742-52.
- HALPERN, Jodi (2012) 'When Concretized Emotion-Belief Complexes Derail Decision-making Capacity', *Bioethics* 26(2): 108–116.
- HOLROYD, Jules (2012) 'Clarifying Capacity: Value and Reasons', in Lubomira Radoilska (ed.), Autonomy and Mental Disorder. Oxford: Oxford University Press, 145–69.
- KORSGAARD, Christine (1989) 'Personal Identity and the Unity of Agency', Philosophy & Public Affairs 18 (2): 101-132.
- MACKENZIE, Robin and WATTS, John (2011) 'Including Emotionality in Tests of Competence: How Does Neurodiversity Affect Measures of Free Will and Agency in Medical Decision-making?', American *Journal of Bioethics Neuroscience* 2(3): 27–36.
- MACLEAN, Alasdair R. (2006) 'Advance Directives, Future Selves and Decision-making', Medical Law Review 14: 291-320.
- MILL, John Stuart (1910/1972) Utilitarianism, On Liberty, and Considerations on Representative Government. London: J. M. Dent & Sons.
- MISCHEL, Walter (2004) 'Integrative Science of the Person', Annual Review of Psychology 55: 1-22.
- NAGEL, Thomas (1970) The Possibility of Altruism. Princeton: Princeton University Press.
- OKAI, David, OWEN, Gareth, MCGUIRE, Hugh, SINGH, SWARAN, CHURCHILL, Rachel and HOTOPF, Matthew (2007) 'Mental Capacity in Psychiatric Patients', British Journal of Psychiatry 191: 291-97.
- PARFIT, Derek (1983) 'Rationality and Time', Proceedings of the Aristotelian Society 84: 47–82.
- PARFIT, Derek (1984) Reasons and Persons. Oxford: Oxford University Press.
- PLANT, Raymond (2011) 'Freedom, Coercion, Necessary Goods and the Rule of Law', Jurisprudence 2(1): 1-16.
- RATCLIFFE, Matthew (2012) 'Varieties of Temporal Experience in Depression', Journal of Medicine and Philosophy 37(2): 114-38.
- RAWLS, John (1971) A Theory of Justice. Cambridge, MA: Harvard University Press.
- REES, Neil (2009) 'Learning from the Past, Looking to the Future: Is Victorian Mental Health Law Ripe for Reform?', Psychiatry, Psychology and Law 16(1): 69-89.
- RUDNICK, Abraham (2002) 'Depression and Competence to Refuse Psychiatric Treatment', Journal of *Medical Ethics* 28: 151–55.
- SCHWARTZ, Robert (1999) 'Autonomy, Futility and the Limits of Medicine', in Helga Kuhse and Peter Singer (eds), *Bioethics: An Anthology*. Oxford: Blackwell, 518–22.
- SJÖSTRAND, Manne and HELGESSON, Gert (2008) 'Coercive Treatment and Autonomy in Psychiatry', Bioethics 22(2): 113-20.

- STRAWSON, Galen (2005) 'Against Narrativity', in Galen Strawson (ed.), Free Will? Oxford: Blackwell Publishing, 63–86.
- SZMUKLER, George and DAWSON, John (2011) 'Reducing Discrimination in Mental Health Law The 'Fusion' of Incapacity and Mental Health Legislation', in Thomas Kallert, Juan Mezzich and John Monahan (eds), Coercive Treatment in Psychiatry: Clinical, Legal and Ethical Aspects. Oxford: Wiley & Sons, 97–119.
- тан, Jacinta, норе, Tony, stewart, Anne and fitzpatrick, Ray (2006) 'Competence to Make Treatment Decisions in Anorexia Nervosa: Thinking Processes and Values', Philosophy, Psychiatry and *Psychology* 13: 267–82.
- woolf, Marie (2005) 'Hockney Blows Smoke on Labour's Plan to Ban Tobacco', The Independent, 29 September 2005.