

Hidden Substance

mental disorder as a challenge to normatively neutral accounts of autonomy

Fabian Freyenhagen and Tom O'Shea,

Essex Autonomy Project,

School of Philosophy and Art History, University of Essex

Mental capacity and autonomy are often understood to be normatively neutral – the only values or other norms they may presuppose are those the assessed person does or would accept. We show how mental disorder threatens normatively neutral accounts of autonomy. These accounts produce false positives, particularly in case of disorders (such as depression, anorexia nervosa, and schizophrenia) that affect evaluative abilities. Two normatively neutral strategies for handling autonomy-undermining disorder are explored and rejected: a blanket exclusion of mental disorder and functional tests requiring consistency, expression of identity, reflective non-alienation, or lack of compulsion. Finally, we suggest ways in which substantivist alternatives to neutrality can be made more promising through increased transparency, democratic contestability of conditions for capacity and autonomy, and a historically sensitive caution concerning restrictions of liberty.

Introduction

The concept of mental capacity continues to influence health and social care practice in many liberal states, and dedicated mental capacity legislation is increasingly common.¹ When

¹ Examples include Scotland's *Adults with Incapacity (Scotland) Act 2000* and England and Wales' *Mental Capacity Act 2005*, as well as narrower guardianship-based legislation centred on mental capacity, such as Ontario's *Substitute Decision Act 1992* and South Australia's *The Guardianship and Administration Act 1993*. The Northern Ireland Executive has proposed to introduce a comprehensive mental capacity bill, and similar legislation is under review in the Republic of Ireland.

someone is found to lack capacity to make a decision, then, in certain circumstances, others are legally permitted to decide on their behalf. This typically includes powers to direct medical treatment, financial affairs, and living arrangements, amongst other matters. Indeed, ‘the right to autonomy’ has even been said to be a person’s ‘full legal capacity to make decisions for themselves’ (Department for Constitutional Affairs 2007:15). Thus, how mental capacity is understood and assessed can be critically important.

Mental capacity and autonomy have contested conditions – each admitting of numerous interpretations – with many accounts aspiring to a form of normative neutrality, thought to befit liberalism. Such neutrality requires states to accommodate pluralism about fundamental ethical, philosophical and religious matters amongst reasonable citizens (i.e. those who themselves respect such pluralism).² Justifications for maintaining neutrality about the good life include claims that this stance is entailed by the moral equality of citizens (Dworkin 1990) and is needed for the legitimation of any coercive political power (Rawls 1996). When extended to mental capacity, normative neutrality suggests that eccentric, unpopular or unwise decision-making should not imply a lack of legal capacity to decide.

Normative neutrality concerning decision-making competence has found considerable legal support. For instance, in English common law, irrationality has been deemed no barrier to competence, even when a decision is ‘so outrageous in its defiance of logic or of accepted moral standards that no sensible person who had applied his mind to the question to be decided could have arrived at it’ (*Re MB [1997]*, para. 30). Furthermore, competent patients have been permitted to refuse even life-saving treatment ‘for any reason, rational or

While capacity and competence are sometimes distinguished – say, by presenting capacity as a psychological term that comes in degrees and competence as a binary legal standard – we will treat them as synonymous for the purpose of this paper.

² We clarify further what is meant by normative neutrality in the next section. Suffice to note already that referring to policies or theories as normatively neutral is not necessarily to say that they are value-free. If there are values and norms that can be endorsed from within different (reasonable) worldviews, then appeal to them is compatible with neutrality.

irrational, or for no reason at all' (para. 16), which implies that being motivated by inconsistent, deplorable or otherwise mistaken value-judgements does not itself preclude competence. Similarly, the *Mental Capacity Act 2005* states: 'A person is not to be treated as unable to make a decision merely because he makes an unwise decision' (§1.4).

We doubt that satisfactory accounts of mental capacity can be given while maintaining normative neutrality. In this article, our primary focus is *philosophical* rather than legal conceptions of personal autonomy – for which mental capacity is a necessary and sometimes also sufficient condition. Specifically, we criticise accounts that deny that autonomy could require evaluative skills or competencies other than those the person being assessed either does or would reasonably accept. Our arguments seek to show that normatively neutral conceptions of autonomy do not (and likely cannot) contain a principled account of mental capacity without smuggling in substantive normative commitments.

Mental disorder is a litmus test which we believe reveals the suppressed normative commitments – the hidden substance – that purportedly neutral accounts of autonomy must rely upon. It is a suitable test for a number of reasons: it is widely agreed (including by proponents of neutral accounts) that mental disorder can imperil mental capacity and thereby individual autonomy; mental disorder also raises issues of liberty, insofar as treatment sometimes involves coercive intervention; and, partly due to longer life expectancy, it is increasingly widespread.

Our aims are mainly negative: to highlight significant shortcomings in purportedly normatively neutral accounts of mental capacity, shown by their inability to identify the capacity-subverting, and so autonomy-undermining, effects of some mental disorders. However, we conclude by suggesting how non-neutral accounts of capacity might exorcise the spectre of unwarranted paternalism, or at least keep it at bay.

Section 1: Normatively neutral conceptions of autonomy

In assessing normatively neutral accounts of autonomy, clarity regarding the kind of neutrality they invoke is crucial. Unfortunately, standardly employed vocabulary and distinctions are problematic; thus, we criticise them, before suggesting a better way to capture the disagreement.

In a legal context, commentators often draw a distinction between outcome-based tests and functional or procedural tests of mental capacity.³ According to the former, persons cannot competently decide to pursue certain outcomes (such as to commit suicide); whereas on the latter, there are no such restrictions, but only a requirement that persons employ certain abilities or deliberative processes in decision-making, or have the capability to do so. Similarly, within philosophical and other scholarly debates, theorists often distinguish between content-neutral and substantive accounts. The latter place *direct* constraints on the content of competent choices, such that only some things can be competently chosen; and the former places no such constraints, or only *indirect* ones – for example, that a person accepts in a minimally rational way their own desires and motives, whatever they are, or would do so if attending to them and their genesis (Christman 1991:22).

While these distinctions are clear enough, they do not adequately locate disagreements about normative neutrality. Crucially, certain conceptions of autonomy contain no direct constraints on the content of choices which can count as competent and also no outcome-based criteria, but their indirect constraints on competent decision-making cannot be adequately understood in a normatively neutral way (Benson 2005:133-5). Instead, these indirect constraints are value-laden – such as requiring autonomous agents to value themselves (see Benson 1994, 2005) – or they rely upon epistemic norms – say, about the degree of warrant required of beliefs that inform the deliberative process (see Wolf 1990).

³ See, for example, Law Commission 1991, esp. para. 2.43-4; see also Owen *et al.* 2009, esp. section 2.

Similarly, drawing the distinction as one between value-neutral and value-laden accounts is imprecise. Firstly, the neutrality in question is not just about values – sometimes the issue is wider and about norms more generally (including epistemic norms governing belief, as just mentioned). Given the nature and degree of pluralism about such norms, proponents of neutral accounts should want to be neutral regarding them too, and not just in relation to any values at stake. Secondly, people – including many without capacity – frequently invoke values in their decision-making, and capacity assessments cannot simply be blind to these values. Indeed, being value-neutral in the sense of discounting every value, including those of the person assessed, could be highly inappropriate and disrespectful. Instead, we must carefully delineate *which* values and norms are at issue: those of the person in question, the assessor(s), their community, or objective values (if there are any). The assessed person's values remain significant even when assessors should be neutral in the sense of not interposing their own commitments or those of the community.

We take neutral accounts to be normatively *internalist* – the only values and norms they subsume within the conditions of autonomy are those of the assessed individual. Internalists exclude other evaluative requirements from assessments of autonomy, such as conformity with social norms or objective values.⁴ For example, inability or systematic failure of a person's decision-making to reflect their own ethical or epistemic commitments could inhibit autonomy on internalist grounds, but only insofar as these commitments are genuinely the person's own. For advocates of normative neutrality, going further and introducing external evaluative criteria, '[...] would make the property of autonomy divergent from the idea of *self*-government that provides its intuitive base' (Christman 1991:14; his emphasis).

⁴ This is not to say that autonomy is all that matters. Someone can be autonomous, but his or her action can still justifiably be restricted if his or her actions harm others in a way that is not compatible with respect for their autonomy and rights (see, for example, Christman 2009 175f n. 26). Our focus is self-harming behaviour, so we shall leave this complication aside.

In contrast, we call ‘substantive accounts’ those that incorporate externalist norms into the conditions of autonomy. These can directly constrain the outcome or content of choices, or instead constrain reasoning or deliberative processes on the basis of competency requirements that make reference to substantial norms and values. Substantive accounts are externalist because the norms which feature as conditions of autonomy do not have to be endorsed by the person whose autonomy is assessed. For instance, substantivists could claim that decision-making competence, and thereby autonomy, requires a person to value themselves in their deliberations, irrespective of that person’s current or former values. Similarly, whether or not people have a commitment to truth, substantive accounts might make autonomy conditional upon acting on true beliefs, or having the capacity to know, and live by, the truth.

Substantive requirements can seem to conflict with an oft-cited fact about modern, liberal-democratic societies: reasonable pluralism about the good, or indeed about other values and norms (including the nature and importance of truth in our lives).⁵ Aversion to substantive and thereby externalist conditions is particularly acute in cases of mental disorder, given what is at stake (such as the potential for long and extensive restrictions of liberty) and the history of abuses in psychiatry. Instead, normatively neutral accounts of autonomy merely impose minimal competency conditions – such as the absence of manifest inconsistency – alongside authenticity conditions – such as coherence of decisions with the person’s overall values and beliefs (see Christman 2011; 2009:ch.7). These conditions, it is proposed, should be

⁵ There is a further complication here: by accepting pluralism only when it is *reasonable*, one moves already away from a purely value internalist picture, for unreasonable persons might not accept that their views are unreasonable and as such subject to interference. It is notoriously difficult to say what ‘reasonable’ means and where its limit should be set, but we take it that for the purposes of liberal political philosophy ‘reasonableness’ consists in accepting that state coercive power should only be used in a way that is compatible with what persons willing to be convinced by arguments can agree on after free discussion. The reasonableness restriction is not always part of accounts of autonomy, such that it remains correct to say that proponents of such accounts aim for normatively neutrality. Still, once such accounts are part of a wider liberal theory, they (tend to) become subject to the reasonableness criterion and the overall theory would only be partially, albeit possibly maximally, neutral as regards to values and norms.

acceptable to each competent person as legitimate constraints. While most adult human beings would count as competent on these conditions – indeed, this is an explicit desideratum of liberal views (see Appelbaum 1998:377) – normatively neutral accounts of autonomy are also meant to identify those genuinely unable to decide for themselves. In the next section, we argue that these accounts are unable to perform this task, particularly when facing individuals with mental disorder.

Section 2: mental disorder as a challenge

Mental disorders can interfere with decision-making capacity and (thereby) autonomy. Of course, not all mental disorders significantly impair capacity and autonomy, nor does everyone with a potentially autonomy-undermining disorder experience the same obstacles to decision-making.

Advocates of substantive and normatively neutral accounts alike agree on the threats posed by certain disorders, such as impairment of ‘cognitive function’ in understanding and recalling information resulting from conditions like vascular dementia. However, when mental disorders are more directly entangled with a person’s evaluative stances, then substantive and neutral approaches suggest divergent accounts of autonomy’s relation to mental disorder. For example, major depression can interfere with decision-making capacity, although not because of any lack of understanding of relevant information, but rather due to stifling negativity or impassive indifference towards future possibilities. Similarly, other mood and anxiety disorders, schizophrenia, anorexia nervosa, personality disorders, and phobias, can all involve implicit normative stances which are not merely unwise, but also

seemingly impede individuals' capacities to make their *own* decisions.⁶ Defenders of normative neutrality aim to avoid making a person's autonomy conditional upon adherence to values or norms which they do not accept, and thus face a challenge: can they account for the capacity-subverting influences of those mental disorders which affect evaluative orientation and judgement without lapsing into a form of substantivism?

We shall consider, and ultimately reject, the two main argumentative strategies open to defenders of neutrality: first, that blanket exclusions for persons affected by mental disorders could be incorporated into the conditions of capacity and autonomy; or, second, that normatively neutral accounts already contain functional or procedural tests which identify capacity-subverting (and so autonomy-undermining) aspects of mental disorder. There is evidence of each strategy within the existing literature, but bringing them to the fore requires reconstructive interpretation, partly because mental disorder is still too often treated as black box – acknowledged as a defeater of autonomy, without any detailed exploration of how, when, and why this is. We focus upon John Christman's influential work, which, rather than being to his discredit, reflects his pioneering attempts to develop a fleshed-out version of normative internalism about autonomy.

2a: blanket exclusion

The blanket exclusion strategy stipulates that mental disorder and autonomy are incompatible: persons with a mental disorder are incompetent and therefore lack autonomy. When articulated so starkly, it remains unclear whether this view finds much support – especially as discussion of mental disorder is so often cursory amongst advocates of

⁶ We consider specific cases below – relying mainly on an appeal to intuition to suggest that autonomy is compromised in these cases.

normative neutrality. Yet, passages which do discuss mental disorder are often beset by a significant ambiguity, which can invite a blanket exclusion reading:

[I] Delusion, paranoia, and other psychopathologies will be inconsistent with autonomy because of the suffering agent's inability to make consistent and reflective judgements about her own set of desires by which she is moved to action. (Christman 1991:17)

[II] [A]utonomy has been conceived as potentially embodying a variety of conditions [... some of which] relate to cognitive and normative competence – rationality, self-control, absence of psychosis and other pathologies, and so on. (Christman 2009:134; see also 135, 147)

[III] The case of a psychotic patient or a senile patient or an infant are also cases where consent is impossible but here it is not a temporary problem. (Dworkin 1988:117)

On a natural interpretation these statements are claims to the effect that competence is simply inconsistent with 'delusion, paranoia, and other psychopathologies', 'psychosis and other pathologies', or being a 'psychotic patient or a senile patient'.⁷

If mental disorders (or some subset thereof) preclude decision-making competence, then there may be no need to introduce externalist norms to identify threats to autonomy posed by disorders that affect evaluative orientation or judgement. Yet, for conceptions of autonomy that exclude mental disorder to remain normatively neutral, the disorders at issue must be specifiable in a normatively internalist fashion. In other words, internalism about autonomy cannot rest on externalism about mental disorder without thereby compromising its neutrality. However, there are good reasons for thinking that an internalist conception of mental disorder is not available.

The concept of *disorder* is normative insofar as its use involves both description and evaluation. To be disordered is not merely to be unusual or a statistical outlier, since it implies failure or incapability relative to norms of good-functioning, such as health or

⁷ However, each statement could be linked to a functional test – the most obvious possibility being a reading of 'because' as 'if and only if' in statement I – and we shall address such interpretations shortly (section 2b).

wellbeing. Thus, employing the concept of mental disorder involves taking a stand on what is valuable, namely how the human mind (and/or brain) ought or ought not to function.⁸ Ordinarily, this judgement is not relative to the normative commitments of the person diagnosed; therefore, it presupposes a normatively externalist perspective, which threatens neutrality. This is most explicit in the influential US American diagnostic manual for mental disorder – DSM-IV – according to which diagnoses of mental disorder are only appropriate if there is ‘clinically significant distress or impairment in social, occupational, or other important areas of functioning’, assessment of which involves consultation of sources ‘in addition to the individual’ (1994:7; see also ICD-10, p. 5). Indexing disorder to occupational and social functioning effectively ties it to historically specific cultural, social, and economic practices, which are deeply saturated by norms and values that the individual might not share.

Commonly-used diagnostic criteria for specific mental disorders also seem to rely upon a normatively externalist perspective, irrespective of whether the very concept of mental disorder does. Consider the ICD-10 criteria for anankastic personality disorder, which include ‘feelings of *excessive* doubt and caution’, ‘*undue* preoccupation with productivity to the exclusion of pleasure and interpersonal relationships’, and ‘*unreasonable* reluctance to allow others to do things’ (F60.5, 1991:205f; emphasis added).⁹ Judging what is excessive, undue or unreasonable presupposes evaluative standards, yet it is characteristic of ‘ego-syntonic’ conditions, such as anankastic personality disorder, that persons diagnosed do not consider their feelings or behaviour to be awry (see DSM-IV 1994:630). Thus, the diagnostic criteria presumably refer to what the majority of psychiatric professionals, or people in general, would take to be excessive, undue or unreasonable (or perhaps to an objective

⁸ The normative nature of mental disorder and diagnostic criteria is contested – we come to this shortly.

⁹ DSM-IV contains a similar characterisation in terms of ‘excessive orderliness’ and ‘inflexibility about matters of morality, ethics, or values’ (1994:629, 669, 670). We return to the issue of inflexibility about matters of morality, ethics, or values below with the example of Martin Luther (see section 2b).

standard for such normative judgements).¹⁰ Moreover, this is not an isolated example: both ICD-10 and DSM-IV littered with normatively charged criteria of an externalist kind – other than ‘undue’, ‘unreasonable’, and ‘excessive’, diagnostic criteria are expressed in terms like ‘inappropriate’ (especially, but not exclusively, compared to ‘age-appropriateness’), ‘irrational’, ‘culturally impossible’, ‘eccentric’, ‘unrealistic’, ‘peculiar’, ‘bizarre’, ‘odd’, ‘unusual’, ‘insensitivity to prevailing social norms/conventions’, ‘inconsistent with subcultural norms’, ‘over-concern/preoccupation’, ‘overvalued idea’, ‘maladaptive [to the social, cultural, and economic environment]’, and ‘expectable’, as well as extra-statistical notions of ‘normal’ and ‘abnormal’. Thus, unless other diagnostic criteria for mental disorder are available, accounts of autonomy which rely upon a blanket exclusion of mental disorder will not be normatively neutral.

The normative basis of psychiatric diagnosis is a highly contentious topic, and some argue that these normative elements can and should be reduced and eventually expelled altogether (for discussion see Bolton 2008; Fulford *et al.* 2005; Sadler 2004; Sadler (ed.) 2002; Wakefield 1992). Therefore, it may be overhasty to criticise attempts at blanket exclusion on these grounds. Nevertheless, the contentious nature of the stand implicitly taken on diagnostic criteria has not even been recognised by those who appear to favour blanket exclusion of mental disorder. Furthermore, such neutral accounts of autonomy stand and fall with the success or failure of normatively neutral conceptions of disorder and diagnostic criteria, and it would be fair to say that the current state of the debate tends more strongly against such conceptions (see Bolton 2008).

¹⁰ There is a tension in the manuals between using cultural and social norms to identify mental disorder and not classifying social critics or dissidents or merely highly eccentric but sane individuals as mentally ill. Both manuals contain clauses to the effect that those who are deviant from or in conflict with society, but where this deviance or conflict is not a symptom of dysfunction, should not be classified as mentally disordered (DSM-IV, 1994:xxif; ICD-10, 1991:5) – presumably partly as a response to the use of psychiatry in suppressing dissidents in the Soviet Union. However, given the fact that being a dissident will often involve social and occupational dysfunction and could be described as ‘maladaptive’ to the social environment, it is difficult to see how this exclusion as specified provides sufficient safeguards. Whilst this tension may be ineliminable, transparency and contestability would be better ways of handling this issue than purported neutrality (see section 3).

Even if these difficulties surrounding the normative substance of psychiatry and medicine can be resolved, the blanket exclusion of mental disorder requires a non-*ad hoc* justification. Yet, beyond a simple correlation between disorder and heteronomy, no such rationale is forthcoming.¹¹ Furthermore, blanket exclusions seem fated to be inapplicable in practice, since many (if not most) assessments of mental capacity require making fine discriminations *within* the population of those already diagnosed with mental disorder. Merely identifying an incompatibility between psychopathology and autonomy provides little help in determining which disorders, decisions and circumstances can authorise paternalistic intervention – especially as current legal regimes, such as the *Mental Capacity Act* in England and Wales, are decision-specific, not status-based, such that a person with a mental disorder might be legally competent to make some decisions (for example, what to wear or eat on a given day) but not others (such as decisions about whether to take their anti-psychotic medication or where to reside).

Finally, defining competence so as to exclude those found to be mentally disordered threatens to be discriminatory. If it is true that people with mental disorder can still make some competent decisions for themselves, then not granting them autonomy rights because of their mental disorder (a status) would discriminate against them compared to the rest of the population who would have these rights denied only on the basis of a decision-specific assessment and incompetence finding. Indeed, the recent UN Convention on the Rights of Persons with Disabilities calls for de-linking of all legislation that can authorise deprivation of liberty (which would include capacity-based legislation, such as the *Mental Capacity Act* in England and Wales), from references to disabilities, be they physical or mental.¹²

¹¹ Alternatively, this could support functional readings, which, through employing specific competence tests, do suggest principled grounds for why certain disorders can impair the evaluative dimension of decision-making (see section 2b).

¹² Thinking of this in terms of ‘de-linking’ is suggested in the UN High Commissioner for Human Rights’s 2009 thematic study on enhancing awareness and understanding of the Convention on the Rights with Persons with

States Parties shall recognize that persons with disabilities enjoy legal capacity on an equal basis with others in all aspects of life.

[...]

the existence of disability shall in no case justify a deprivation of liberty. (Art. 12(2) and Art. 14(1b))

Mental capacity assessment that proceeded solely on the basis of establishing whether or not the person has a mental disorder, and which subsequently led to deprivations of liberty solely on the basis of a finding of mental disorder, would thus contravene the Convention and, at any rate, seems discriminatory.¹³

2b: functional tests

The second strategy employed by proponents of normatively neutral conceptions of autonomy is to offer functional tests that are meant to identify autonomy-subverting influences without incurring a commitment to specific values, substantive notions of rationality, and the like. This chimes well with legal provisions of liberal-democratic societies, in which functional tests of competence or capacity are often operative. For example, in England and Wales, the decision-specific test of the *Mental Capacity Act* ultimately premises mental capacity on the ability to understand relevant information, to retain it long enough to make a decision, to use or weigh it, and to communicate the decision formed, while recognising that possessing these abilities need not imply that one avoids making unwise decisions (or what others would judge to be such). Similarly, in many US jurisdictions, competence rests on the capacity to understand the relevant information,

Disabilities, para. 48-9 (see <http://www2.ohchr.org/english/bodies/hrcouncil/docs/10session/A.HRC.10.48.pdf>, last accessed 27/06/2011). We thank Phil Fennel for making us aware of this source. Other interpretations of the Convention are more radical and take it to oppose even functionalist models of legal capacity (see Dhanda 2007).

¹³ There is one further worry about the blanket exclusion strategy: given the prevalence of mental disorder, it runs the risk of going against the liberal desideratum of counting most adults among the competent.

appreciate how it applies to one's own case, reason on the basis of it, and then express a choice, but there is no (direct) constraint on the content of decisions or values that can be competently adopted. It is an interesting question whether the legal tests are indeed normatively neutral both in spirit and practice (see Owen *et al.* 2009, especially section 2; Richardson 2010), but we will concentrate here – at least in the first instance – on the philosophical models.

Specifically, we will consider the following four functional tests (here formulated as questions, such that negative answers to them indicate lack of capacity and autonomy):

- A. Is the decision minimally rational in the sense that it does not proceed from a manifestly inconsistent set of desires or beliefs?
- B. Does the decision – or the preferences and values on which it is based – express the person's enduring identity?
- C. Is the decision based on preferences and values from which the agent would not be alienated, if she were to reflect on their genesis?
- D. Is the decision made without compulsion, such that the person could choose and act otherwise?

These present the most promising attempts to formulate neutral tests of autonomy, but – as we shall argue – they are, nonetheless, inadequate.¹⁴

A. As mentioned earlier, normatively neutral accounts typically contain competency conditions, including a requirement of minimal rationality of the decision-making process,

¹⁴ Frankfurt's structural conception of autonomy is not discussed here, but would also fail to account for the cases we consider below. Specifically, persons with autonomy-undermining, but ego-syntonic mental disorders can be said to identify with their lower order volitions at higher orders – they are like Frankfurt's willing addicts (see 1971), but nonetheless (can) lack autonomy.

The *Mental Capacity Act's* test (understanding, retaining, using or weighing decision-relevant information and expression a choice) is also not discussed, partly because the proposed four tests considered here could be understood as offering philosophical explorations of its crucial, but little understood, 'use or weigh' dimension (even though they were not expressively meant as such by the authors in question).

whether in relation to an actual process of reasoning, or hypothetically, such that this reasoning could have been successfully undertaken (see Christman 1991:11, 16f, 22; 1993:282f; 1995:35; 2009:134, 154f; see also Arneson 1994:47, 61). In Christman's influential account, this is spelled out as the requirement that the (actual or hypothetical) reasoning process does not involve '*manifestly inconsistent* desires or beliefs' (1991:15, Christman's emphasis; see also 13fn22, 16, 17; 1993:287, 288; 2005:278; 2009:155; see also Berofski 1995:10):

Those preferences or beliefs that are in obvious conflict, ones which the agent could bring easily to consciousness and recognize as incompatible, are what one would label "manifestly inconsistent". (Christman 1991:15; see also 1993:287)

If agents have manifestly inconsistent desires or beliefs, they cannot be said to be self-governing, for they are then mere playthings of conflicting forces.¹⁵ Also, failing to meet even such a minimal rational requirement suggests that the person is unable to engage in the critical reasoning required by normatively neutral accounts of autonomy (see Arneson 1994:46f). Still, this rationality requirement is relatively lax, since few of us have fully consistent desires and beliefs, and it is one of the key intuitions of liberal accounts of autonomy that most adults should be found to possess the capacity in question (and thus the moral status that underpins liberal rights and protections against the state and others; see Appelbaum 1998:377; Feinberg 1989:28).

However, it is not convincing to take absence of manifest inconsistencies as sufficient to exclude the autonomy-subverting influence of mental disorder. While it might be true of some psychopathologies associated with deficits of autonomy that those suffering from them

¹⁵ Is requiring an absence of minimal inconsistency an internalist norm (that is, one that applies whether or not each person to whom this norm is applied accepts it)? Sometimes, it sounds as if it is meant to be a (quasi)transcendental condition of competent agency (hence the formulation used in the main text), and if so, it would be internalist by default.

are subject to manifest inconsistencies, this is by no means true of all of them. Consider two cases:

Anorexic Lucy: Lucy has been diagnosed with anorexia nervosa for several years, and her body weight is now critical. Unless she gains weight very soon, there is a significant risk that she will suffer organ failure and death. Although sometimes finding it hard to concentrate and think, she is usually lucid and can pass various cognitive tests. Whilst not actively wishing to die, Lucy repeatedly states that she still feels too fat, takes herself to be sufficiently well to doubt that her condition is as critical as the doctors say, and, besides, is more averse to gaining weight than dying. For Lucy, the commitment to thinness and bodily control overshadows all other values and relationships, and has become integral to her sense of self.¹⁶

Deluded Roger: Roger has been diagnosed with schizophrenia. He has a particular delusion, thinking himself to be God. The doctors claim that he suffers from diabetes, but he refuses to take any medicine for it, even after being warned that it will endanger his health, well-being, and eventually his life. While he accepts that people who suffer from diabetes should take medication, he reasons that as an immortal, all-knowing and all-powerful God, he could not be affected by, never mind die from, a physical disease and, at any rate, would know if the material body in which he appears to us were ill, and could cure it by an act of will.¹⁷

It would seem that both of these individuals lack the decision-making capacity necessary for autonomy, but neither of them is subject to manifest inconsistencies. If anything, they are models of consistency – with a fully worked-out hierarchical value system (Lucy) and belief system (Roger). The reason why we would still say that they lack capacity and autonomy (at

¹⁶ While this case is fictional, its components build on the experiences and views reported by persons with anorexia nervosa and their mothers in Tan *et al.* 2006. One of these persons – Participant I – states: ‘I wasn’t really bothered about dying, as long as I died thin’ (*Ibid.*, 274; see also 275); and insists that anorexia nervosa is part of her identity (*Ibid.*, 276).

¹⁷ This example derives from real life cases, knowledge about which we owe to a Wellcome Trust funded study by Gareth Owen on decision-making in three groups of patients (those with severe depression, schizophrenia, or frontal lobe damage). We thank him for access to this material.

least in respect to the respective decision in question) is better cashed out in terms of substantive conceptions of autonomy: both Lucy's decision to refuse food – due to her valuation of thinness and control above all else – and Roger's to refuse diabetes medication – because of a delusional belief that he is God – are expressing, for example, an inability 'to cognitively and normatively understand and appreciate the world for what it is' (Wolf 1989:150).¹⁸

Substantive accounts, while capturing (better) why competence or capacity is missing, eschew normative neutrality in virtue of their ethical and metaethical commitments (see also Wolf 1990). Despite conflicting with legal rhetoric, such accounts actually fit well with legal practice. Thus, the authors of the MacArthur Competence Assessment Test for Treatment (MacCAT-T), an assessment tool distilled from US common law, when commenting on cases of patients with *anorexia nervosa* like Lucy, write that their ability to appreciate is impaired here because 'the degree to which they valued being thin (or detested being fat) involved a considerable distortion of reality', not least in thinking that 'nothing in the world – friends, family, life itself – is of any value except thinness' (Grisso and Appelbaum 2006:296; see also Grisso and Appelbaum 1998:45). Insofar as this involves a clear value judgement (specifically, that thinness is not the only or paramount value) and their views really express the common law tradition in the USA, then legal practice there is in fact value-laden in an externalist way. Similarly, recent UK legal cases show that judges are prepared to rule that persons lack capacity if they do not appreciate the seriousness of their condition and if their decision is based on a misperception or misconception of reality (see, for example, *Re T [2004]*; *Trust A and Trust B v. H [2006]*). This suggests that respect for decisions based on

¹⁸ Wolf speaks of freedom, not autonomy – rejecting the latter because she equates it with metaphysical self-creation, the possibility of which she denies. Since ascription of autonomy is not commonly seen to include the metaphysical thesis Wolf rejects and the conditions she proposes can be conceptualised well in terms of self-determination (not just freedom), it is standard practice to speak of Wolf's view as one about autonomy, and we follow this practice here.

‘irrational reasons’ is in legal practice much more circumscribed and evidences a more substantive approach than the legal rhetoric suggests.¹⁹ The more substantial approach is even clearer when we look beyond mental capacity to legislation on involuntary hospitalisation because of risks to oneself (and/or others), such as the *Mental Health Act 1983*, amended 2007, in England and Wales.

Be that as it may, it is important to distinguish the consistency of desires, values and beliefs, as well as consistent reasoning on the basis of them, on the one hand, from their truth, justification, or coherence with well-supported ethical outlooks and scientific conceptions of the world, on the other (see also Bayne and Levy 2005:80). The former are achieved by many of those with mental disorders – indeed, it has been suggested that those with certain mental disorders reason more logically than other people (see Owen, Cutting and David 2007). Whatever is problematic about having delusional beliefs, it is naïve to think that possessing them means that one would necessarily fail the minimal inconsistency test. Arguably, consistency is a prerequisite of a minimum psychic order necessary for possessing the capacity for autonomy (see Arneson 1994:55; Oshana 1998:94), but both Lucy and Roger, and many others whose capacity and autonomy are in doubt, meet this criterion.

B. Other elements of purportedly normatively neutral functional tests are also problematic. Thus, it is sometimes proposed that it is a condition of competence and thereby autonomy that the preferences and values that guide someone’s life be part of his or her enduring identity (Christman 2009:149; Craigie 2011; see already Christman 1991:15). However, such a test is unsuited to handle autonomy-subverting influences. As we saw earlier, *Anorexic Lucy* has fully integrated her condition and value orientation into her

¹⁹ Indeed, even the liberally neutral sounding rhetorical flourish in *Re MB* [1997] – quoted in the introduction -- did not stop the judges from finding the person before the court to lack capacity. With few exceptions, this is common pattern.

identity to the extent that if someone could remove it by – say – waiving a magic wand, she would deny that the person emerging from this would be herself.²⁰ Again, many mental disorders that undermine autonomy involve conflicted identities and/or self-alienation (such as obsessive compulsive disorder or borderline personality disorder), and this might suggest that people with these disorders lack an enduring identity and, in virtue of this, they lack autonomy. Yet, not all disorders do involve such conflicted identities or self-alienation – indeed, recall that not even all personality disorders do, since persons with ego-syntonic conditions, such as anankastic or anti-social personality disorders, typically have non-conflicted identities and stable beliefs and desires. We would still want to say that autonomy is undermined by many such conditions, including those of *Anorexic Lucy* and *Deluded Roger*.

C. Even Christman’s otherwise laudable move to include the history of the belief- and preference-formation into the conception of autonomy does not suffice to exclude autonomy-subverting factors. He proposes that autonomy requires that were the individual to become aware of the history of their belief and motivation set, then they would not be alienated from this set. Again, this is meant as a relatively minimal condition: neither actual reflection is required (merely non-alienation if the agent were to critically reflect), nor is the standard of transparency required for such reflection set at a demandingly high level (Christman 1991:11f; 2009:154). While this has the advantage of not excluding the bulk of the population, it creates false positives for many people suffering from mental disorder. Thus, *Anorexic Lucy* might discover, say, by way of psychotherapy, that her drive to be thin at all costs is a reaction to overbearing parenting. She might also learn from feminist theories and sociology more generally that a fixation on thinness is often fostered by social expectations of

²⁰ See Tan *et al.* 2006:276.

women, including contentious cultural norms regarding demeanour, femininity and perfectionism, which are expressive of a deeply patriarchal society. Nevertheless, Lucy might respond with indifference or even gratitude in discovering this, rather than revising her commitments. Perhaps she views the overbearingness of her parents as a reason to continue to assert her will in the one area where it is very difficult to wrestle control from her (her body), and decides to appropriate and subvert, rather than reject, the gender identity thrust upon women by society. Similarly, someone might recognise that certain self-harming behaviours (cutting or refusal of food) are the result of an abusive childhood, but view this as her only real way of coping, and accept it along with the possible consequences of death or severely reduced quality of life.²¹ If proponents of normative neutrality want to accept that persons can autonomously endorse unusual, even bizarre preferences – as Christman does (1991:21n32, 23) – then they lack the resources to account for the lack of capacity (and thereby autonomy) in these examples.

D. Finally, advocates of neutral accounts could also appeal to the compulsive nature of decisions made by those with certain mental disorders. For example, in the Code of Practice to the *Mental Capacity Act* (§4.22), it is said that a person with anorexia nervosa may lack capacity in relation to some decision because ‘their compulsion not to eat might be too strong for them to ignore’, rendering them unable to use or weigh relevant information (even if it is understood). This approach to anorexia is traceable to an earlier judgement in the Court of Appeal (*Re W* [1992] 4 All ER 627 at 637 and 639), in which Lord Donaldson argues:

[...] it is a feature of anorexia nervosa that it is capable of destroying the ability to make an informed choice. It creates a compulsion to refuse treatment or only to accept treatment

²¹ This latter scenario is also based on a real case: *LB v. Croydon Health Authority* [1995] 1 FCR 332; and *B v. Croydon Health Authority* [1995] Fam. 133 (Court of Appeal).

which is likely to be ineffective. This attitude is part and parcel of the disease and the more advanced the illness, the more compelling it may become.

No direct objection is raised to the values of the anorexic person – instead, mental capacity is meant to be impaired by their inability to utilise information in decision-making.²² Thus, more generally, if compulsiveness is an effect of those mental disorders that appear to undermine evaluative capacities, then sufficient psychological pressure, such that someone cannot think, decide or act otherwise, might explain failures of self-determination without the need to invoke normatively externalist criteria.

However, autonomy is not always impeded by compulsion, even in circumstances in which an individual cannot do otherwise. Consider the apocryphal words that Martin Luther is said to have spoken when defending his writings at the Diet of Worms: ‘Here I stand; I can do no other’ (see Dennett 1984:133 and Wilson 2007:153, 169-70). If Luther genuinely lacked the psychological capability to renounce his works – or his acting against conscience would have been spiritually or rationally intolerable – then were his actions thereby less autonomous? Quite the reverse – it would seem that weaker convictions, which are not too strong to ignore, are less likely to be expressions of what a person takes to be fundamentally valuable or important, and therefore can provide a correspondingly less exemplary basis for self-determined action. Similarly, someone identifying as gay, or a dissident accused of ‘delusions of reformism’, is likely to feel compelled to refuse purported treatments, where, so to speak, ‘this attitude is part and parcel [of their condition]’ and ‘the more advanced [it is], the more compelling it may become’, yet this casts no doubt on their mental capacity.²³ In the

²² The legal exclusion of compulsiveness is mirrored by philosophical demands for value-flexibility. For example, Dworkin requires there be ‘[...] some ability both to alter one’s preferences and to make them effective in one’s actions’ (1988:17).

²³ In line with our argument, Christman denies that autonomy requires an ability to change our fundamental commitments, or even that we can reflect separately from them (2009: 121-32). He only accepts a more qualified requirement, namely, the ability to stand back from these commitments *only when and because* we are deeply alienated from them. However, neither Luther nor *Anorexic Lucy* is deeply alienated from their respective commitments.

absence of non-neutral normative criteria which would identify pathological kinds of compulsion, then compulsiveness alone is not sufficient to impair autonomy. Consequently, *Anorexic Lucy* and her decision to refuse force-feeding cannot be counted as incompetent by advocates of normative neutrality on these grounds, even if there is a sense in which her condition compels her decision.

None of the proposed functional tests adequately capture the capacity-subverting, autonomy-undermining aspects of mental disorder. Each test has been shown to fail in isolation and when applied jointly (as case of *Anorexic Lucy* that ran through the discussion of all of them demonstrates). We should, however, discuss a possible response by a defender of normative neutrality. Our argument depends upon the identification of various false positives: persons lacking autonomy who cannot be identified as so lacking by normatively neutral accounts. However, advocates of normative neutrality might bite the bullet here – acknowledging that counting persons such as Lucy or Roger as autonomous is counterintuitive, but accepting this as a revisionary consequence of their view.

In response, we can ask why should self-harming, ill-informed or irrational decisions be respected? Mill might be thought to provide an anti-paternalist justification.²⁴ He writes:

[...] the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right. (Mill [1859]/1991:14).

²⁴ Indeed, opponents of compulsory treatment for those with mental disorder have appealed to Millian considerations in support of their view (see, for example, Szasz 1970:116).

The reasons he offers are that individuals are best-placed to judge their own interests, that the moral equality of persons demands respect for others' liberty, and that paternalism would inhibit character-development.

However, Mill makes an important qualification:

It is, perhaps, hardly necessary to say that this doctrine is meant to apply only to human beings in the maturity of their faculties. [...] Those who are still in a state to require being taken care of by others, must be protected against their own actions as well as against external injury. (Mill [1859]/1991:14).

Thus, appealing to Mill's harm principle raises the further question of under what circumstances someone is 'in the maturity of their faculties' – precisely what is at issue regarding Lucy and Roger. Mill's examples of immaturity are developmental: young people before reaching adulthood and, with an odious ethnocentrism, 'barbarians' from 'backward states of society' who are not yet 'capable of being improved by free and equal discussion' (Mill [1859]/1991:14). However, the salient traits of maturity, including a capability to know one's own interests, can also presumably be lost, such as through dementia and (as we hope to have made plausible) through temporary or permanent failures of evaluative abilities wreaked by mental disorder. In light of this justification for restricting liberty, Mill may not prove to be an unalloyed ally of normative neutrality.²⁵

Even were a more radical line taken, disregarding Mill's qualifications, a problem remains: it is not neutral to champion a form of liberty whose only bounds are harm to others. Some dignitarians, for example, claim that human dignity can take precedence over individual liberty – an argument that, for example, France's *Conseil d'Etat* endorsed considering 'dwarf tossing' (27 Octobre 1995, N° 136727, *Commune de Morsang-sur-Orge*),

²⁵ Also, Mill is, in fact, willing to place direct constraints on what counts as autonomous choices: entering into any irrevocable contract and especially selling or giving oneself into slavery cannot be done autonomously (and does not entitle the person a right against paternalistic interference), for '[t]he principle of freedom cannot require that he should be free not to be free. It is not freedom to be allowed to alienate his freedom' (Mill, [1859]/1991:114).

subsequently upheld by the United Nations Human Rights Council in *Wackenheim v. France*. Other dignitarians argue that respect for self-determination cannot entail the right to make decisions which conflict with dignity (such as some decisions to commit suicide), since human dignity is the very basis of this respect (Velleman 1999). Notably, whether or not there is a conflict with dignity is not up to the individual alone on this view, but is decided by public, intersubjective reasoning. We do not mean to endorse the dignitarian view, but merely use it to make a general point: even if actions, decisions, or people falling under the scope of the harm principle can be individuated in a normatively neutral fashion, whether the principle is justified remains deeply contested in a fashion precluding neutrality.

Section 3: replacing neutrality with transparency and accountability

We have shown that attempts to develop normatively neutral approaches to autonomy are unsuccessful: either they produce false positives (thereby failing as criteria of autonomy) or they rely upon hidden and contentious substantial commitments (thereby forsaking neutrality). It now looks as if we face an impasse, since the alternative — substantivism — has been thought to be unduly paternalistic, even authoritarian, and to open the door to abuses of power. Nonetheless, ours is not a counsel of despair. Instead, we suggest practical steps to rehabilitate substantivism.

Substantivism can seem most unattractive in its strongest forms, which place direct constraints on what people may autonomously decide or the reasons for their decisions. However, substantivism can also take a different form – indirect or, as it is also known, weak substantivism – and adopting this kind might begin to allay the concerns. For instance, Paul Benson advocates a normative-competence account, on which autonomy ‘demand[s] that agents’ capabilities of perception, reasoning and motivation be connected in the right sorts of

way to what is really valuable or reasonable for them' (2005:134). Since what is really valuable or reasonable for people need not be relative to what they take to be valuable or reasonable, then this account is substantive. However, formulating conditions for autonomy in terms of agents' capabilities (rather than the content, motivation, or propriety of individual decisions) makes room for people to autonomously make individual unwise decisions, so long as there is not a systematic distortion in their evaluative skills.²⁶

Our core examples of schizophrenic delusion and anorexia nervosa are plausibly cases of such systematic distortion, where we can identify recurrent failures of perception, reasoning or motivation which prevent people deciding with their genuine interests properly in view. Thus, indirect substantivism would not fall prey to the same false positives as normatively neutral accounts of autonomy, yet promises to be less prescriptive than direct substantivism. In emphasising abilities and skills, there is also a focus on functions instead of status, thereby minimising the discrimination that status-based approaches to capacity are often thought to imply. The challenges facing indirect substantivism remain considerable, and we do not suggest dangers like tyranny of the majority, unjustified paternalism, and abuse of power, can be completely put to bed; but we also believe that the situation is not hopeless. In particular, we think substantivism can be buttressed by *increased transparency*, the *democratic contestability* of conditions for capacity and autonomy, and a *historically sensitive caution* concerning restrictions of liberty.

When an account of autonomy is meant to inform a conception of mental capacity, and therefore people's liberty is at stake, then transparency concerning its conditions is essential. Imposing potentially contentious externalist conditions – presupposing beliefs and values that others may not share – makes this clarity especially important. Arguably, even

²⁶ Wolf's position (invoked above in section 2b) might also be understood as indirectly substantive because it simply requires '[...] the *ability* to cognitively and normatively understand and appreciate the world for what it is' (1989:150; Wolf's emphasis).

some existing legislation and guidance falls short in this respect – for instance, disagreement persists about the legitimacy of invoking external norms when applying the *Mental Capacity Act* test that someone be able to ‘use or weigh’ information relevant to the decision to be made (§3(1)(c); see also Tan *et al.* 2009). The main danger of such opacity is arbitrariness and undue discretionary authority in mental capacity assessment, which places assessors in positions of dominance, threatening the assessed individual’s freedom, even if assessors exercise their powers responsibly.²⁷

Despite seeming benign, greater transparency concerning the norms underlying capacity assessment might be resisted on the grounds that these standards can be handled implicitly (as in much physical diagnosis) and that demanding too much explicitness risks limiting the flexibility available to assessors to implement pragmatic solutions on a case-by-case basis.²⁸ The concern is that difficult-to-codify assessment skills that come with experience become marginalised in favour of clear but crude explicit principles, leading to both false positive and negative ascriptions of mental capacity.

We take these concerns seriously and would like to see training that develops skills which good assessors can exercise confidently (with appropriate institutional oversight) rather than requiring continual recourse to strict principles that increase inter-rater reliability only at the price of the validity of assessments (see also Fulford *et al.* 2005:80f). However, increased transparency does not preclude this. When the values and beliefs guiding assessment are explicitly thematised, this allows greater scrutiny, opening them to revision in

²⁷ On freedom as non-domination, see Pettit 1997: ch.2-3 and 2001:138-51. If we look to the history of psychiatry, we find that moves to require explicit discussion and recording of what were formerly implicitly justified practices (such as mechanical constraint of psychiatric patients) were an important check on the arbitrary use of power (Fennell 1996).

²⁸ Furthermore, it might be thought to constrain *judicial* pragmatism (for good or ill), since, in practice English courts have afforded themselves generous scope to disregard decisions as incompetent – often, in effect, exercising their own values – particularly when it comes to refusal of treatment (see Richardson 2010).

light of reasoned criticism.²⁹ We acknowledge that greater transparency need not necessarily lead to increased or better consensus. Sometimes, it can produce more strife and entrench opposition. Our plea for transparency is not based on the judgement that there are no risks attached to it, but merely that there are greater risks in not taking this path – namely, that substantive and supposedly normatively neutral approaches alike smuggle in contentious evaluative commitments without discussion.

Secondly, alongside transparency about their presence, substantive evaluative conditions of mental capacity should be subject to democratic accountability and contestability. Transparency opens up the terrain of value; then, what evaluative capabilities are relevant can be settled by political discussion and argument (see also Bolton 2008: 224, 238f). In order to forestall majoritarianism, such that values and beliefs are not imposed on others with little or no recourse to the voices of those affected by such democratic decisions, then this may need to be an asymmetric kind of debate – more weight being given to the views of those who have or have had mental disorders, as well as those involved in supporting and caring for them, such as psychiatrists and social workers. In addition, the dangers of holding others to merely parochial standards could be ameliorated by attempts to achieve an ‘open impartiality’ (Sen 2009: ch.6) that assimilates a maximally heterogeneous set of cultural perspectives and identities. In practice, this also signals the importance of a diverse workforce of those charged with making mental capacity assessments, so that the values and beliefs of some groups (e.g. white middle-class men) do not predominate uncontested.

We think – not without controversy – that capacity assessment, based on norms not shared by the assessed person, and leading to restriction of liberty, is sometimes necessary;

²⁹ Indeed, it has been argued that psychiatry’s vulnerability to abuse for political purposes in the Soviet Union was due to ‘[...] a failure to recognise the extent to which the Soviet values were influencing clinical judgements’ (Fulford *et al.* 2005:83, with references). This suggests that the transparency we are calling for might make psychiatry less vulnerable to such abuse.

but the harms of these interventions can be lessened by being shaped by those who have experienced being subject to them (and subjecting others to them). For instance, this might involve a process of institutional oversight – involving those with mental disorders, people close to them, care workers, psychiatrists, and so on – guided by such factors as the views of former patients concerning whether and when they would have wanted another decision-maker to take control of decisions around treatment or financial affairs, and the like. Again, this process should follow the principle that those affected by requirements for evaluative competencies should always have a collective voice in deciding how these conditions are introduced and interpreted, even if in specific cases an individual's wishes cannot always be fully respected.

Finally, let us comment further on worries about abuses of power by care workers and mental health professionals. In psychiatry, substantive diagnostic conditions have a particularly dark history – recall the examples of hospitalisation for homosexuality and ‘delusions of reformism’ mentioned above. However, not all substantive conditions should be rejected on this basis. Instead, this provides a reason to implement what we call a ‘*negative capacity regime*’. Such a regime would be founded upon a ‘concrete negation of historical wrongs’ (Niesen 2004:83) and be oriented by the need to avoid repetition of past abuses. In this respect, historically informed caution concerning the harms of abuse (rather than respect for autonomy *per se*) speaks in favour of a very high threshold for non-consensual intervention even when mental capacity is found to be lacking.

Once more, this is not a silver bullet – difficult questions remain: what should the exact lessons be from past abuses? How and when is a treatment sufficiently similar to the past abuses, despite other differences, to trigger heightened concern and safeguards? How can we avoid overcompensating for past abuses by letting too many false positives slip through

our anti-interventionist net? What we do say is that practices are likely to be better for keeping the past in view, not that there no dangers down this path.

By way of conclusion, let us sketch briefly what our approach to mental capacity would mean in practice.³⁰ We already mentioned diversification of the workforce. Alongside this, training of psychiatrists and other capacity assessors would have to change to emphasise the value- and norm-laden nature of disorder, diagnostic criteria, and the capacity assessment process (some of the required training material is already being developed under the heading of ‘value-based practice’; see Fulford *et al.* 2005:80f, 83). Strategies for navigating the challenges involved in fostering normative competencies, whilst allowing unwise decisions when these competencies are in place, should also feature in training. Changes in psychiatric practice might also be necessitated – notably, to encourage more engagement with the content of people’s beliefs and values as well as how they experience them – and in the way capacity assessment and best-interest decisions are recorded and made. Democratic mechanisms would have to establish and regularly review what normative competencies should be considered as requisites of autonomy and the legal rights that come with it. At least the broad outlines of what should count as (potentially autonomy-undermining) mental disorder may have to be decided by democratic mechanisms of a society as a whole, not just by specialist bodies. Organisations of people with the conditions said to undermine these competencies, and those close to or caring for them, would need to have a major role in these democratic mechanisms, allowing ample opportunity for contestation. Also, those who have experienced life with these conditions might have to be included as lay members onto case review panels – just as already mental health tribunals often include lay members. Ombudsmen and advocate services should also become more frequent. Judges and other key decision-makers

³⁰ Many of the measures below would have resource implications, but given the high stakes involved (such as deprivations of liberty), liberal societies have to shoulder them if they are to live up to their ideals, just as compromising due process would be unjustified, even if it saved costs.

would have to regularly engage the public in such a way as to make the decisions in question maximally transparent to lay persons. Finally, every effort should be made to avoid interference with liberty – often creative thinking about treatment or care provisions can help here instead of accepting a narrow set of alternatives (such as treatment against refusal or no treatment at all) – and, where non-consensual intervention is unavoidable, there have to be robust safeguards.

None of this will provide cast-iron guarantees against unwarranted paternalism, and significant obstacles to indirect substantivism remain. Yet, we believe that its prospects are brighter than purported normative neutrality, which – if the arguments in this paper are sound – is either toothless or hides its substantive elements, rather than submitting them to public scrutiny.³¹

³¹ For comments on and criticisms of earlier drafts, we would like to thank Viv Ashley, Jill Craigie, Wayne Martin, Gareth Owen, Lucy Series, and Antal Szerletics (who also provided research assistance) as well as two anonymous referees of this journal.

Bibliography

- Adults with Incapacity (Scotland) Act 2000*.
- Mental Capacity Act 2005* [England and Wales].
- Mental Health Act 1983*, amended 2007 [England and Wales].
- Substitute Decision Act 1992* [Ontario, Canada].
- The Guardianship and Administration Act 1993* [South Australia, Australia].
- Re W [1992] 4 All ER 627.
- LB v. Croydon Health Authority [1995] 1 FCR 332;
- B v. Croydon Health Authority [1995] Fam. 133 (Court of Appeal).
- Re MB [1997] EWCA Civ 3093.
- Re T [2004] EWHC 1279 (Fam.)
- Trust A and Trust B v. H [2006] EWHC 1230 (Fam.)
- Re P [2009] EWHC 163 (Ch.)
- Conseil d'Etat*, 27 Octobre 1995, N° 136727, *Commune de Morsang-sur-Orge*.
- Wackenheim v. France* (2002, UNHRC Comm. No. 854/1999, CCPR/C/75/D/854/1999).
- Diagnostic and Statistical Manual of Mental Disorder* 4th Edition [DSM-IV], Washington, D.C.: American Psychiatric Association, 1994.
- The ICD-10 Classification of Mental and Behavioural Disorder: Clinical Descriptions and Diagnostic Guidelines* [ICD-10], Geneva: World Health Organisation, 1991.
- Appelbaum, Paul S. (1998) 'Ought We to Require Emotional Capacity as Part of Decisional Competence?', *Kennedy Institute of Ethics Journal* 8.4: 377-87.
- Arneson, Richard J. (1994) 'Autonomy and preference formation' in J.L. Coleman & A. Buchanan (eds.) *In Harm's Way: Essays in Honor of Joel Feinberg*. Cambridge: Cambridge University Press, Ch. 3.
- Bayne, Tim and Levy, Neil (2005) 'Amputees By Choice: Body Integrity Identity Disorder and the Ethics of Amputation', *Journal of Applied Philosophy* 22.1: 75-86.
- Benson, Paul (1994) 'Free Agency and Self-Worth', *The Journal of Philosophy* 91.12: 650-68.
- _____ (2005) 'Feminist Intuitions and the Normative Substance of Autonomy', in J.S. Taylor (ed.), *Personal Autonomy*. Cambridge: Cambridge University Press, Ch. 5.
- Berofsky, Bernard (1995) *Liberation from self: a theory of personal autonomy*. Cambridge: Cambridge University Press.
- Bolton, Derek (2008) *What is mental disorder? An essay in philosophy, science, and values*. Oxford: Oxford University Press.
- Christman, John (1991) 'Autonomy and Personal History', *Canadian Journal of Philosophy* 21.1: 1-24.
- _____ (1993) 'Defending Historical Autonomy: A Reply to Professor Mele', *Canadian Journal of Philosophy* 23.2: 281-290.
- _____ (1995) 'Feminism and Autonomy', in D.E. Bushnell (ed.), *"Nagging" Questions: Feminist Ethics in Everyday Life*. Lanham, Maryland (etc.): Rowen & Littlefield, Ch. 2.
- _____ (2004) 'Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves', *Philosophical Studies* 117: 143-64.
- _____ (2005) 'Procedural Autonomy and Liberal Legitimacy', in J.S. Taylor (ed.) (1995), *Personal Autonomy*. Cambridge: Cambridge University Press, Ch. 12.
- _____ (2009) *The Politics of Persons: Individual Autonomy and Socio-historical Selves*. Cambridge: Cambridge University Press.

- _____ (2011) 'Autonomy in Moral and Political Philosophy', Edward N. Zalta(ed.), *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, first published 2003, <<http://plato.stanford.edu/archives/spr2011/entries/autonomy-moral/>>.
- _____ (ed.) (1989) *The Inner Citadel: Essays on Individual Autonomy*. Oxford: Oxford University Press.
- Craigie, Jill (2011) 'Competence, Practical Rationality and What a Patient Values', *Bioethics* 25.6:326-33.
- Dennett, Daniel (1984) *Elbow Room: The Varieties of Free Will Worth Wanting*, Oxford: Oxford University Press.
- Department of Constitutional Affairs (2007), *Mental Capacity Act 2005 Code of Practice*. London: TSO [The Stationery Office].
- Dhanda, Amita (2007) 'Legal Capacity in the Disability Rights Convention: Stranglehold of the Past or Lodestar for the Future?', *Syracuse Journal of International Law and Commerce* 34.2: 429-62.
- Dworkin, Gerard (1988) *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.
- Dworkin, Ronald (1990) 'Foundations of Liberal Equality', in *Tanner Lectures On Human Value Vol XI*, ed. by Grethe B. Peterson. Salt Lake City, UT: University of Utah Press, pp. 1-119.
- Feinberg, J. (1989), 'Autonomy', in Christman (ed.) 1989: Ch. 2; reprinted from his *Harm to Self*, Oxford: Oxford University Press, 1986, Ch. 18.
- Fennell, Phil (1996) *Treatment without consent: law, psychiatry and the treatment of mentally disordered people since 1845*. London: Routledge.
- Frankfurt, Harry (1971) 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy* 68.1: 5-20.
- Fulford, K.W.M. *et al.* (2005) 'Looking with both eyes open: fact and value in psychiatric diagnosis?', *World Psychiatry* 4.2: 78-86.
- Grisso, Thomas and Appelbaum, Paul S. (1998) *Assessing Competence to Consent to Treatment: A Guide for Physicians and Other Health Professionals*. New York: Oxford University Press.
- _____ (2006) 'Appreciating Anorexia: Decisional Capacity and the Role of Values', *Philosophy, Psychiatry, & Psychology* 13.4: 293-7.
- Law Commission [England and Wales], *Consultation Paper No. 119* (1991): Mentally Incapacitated Adults and Decision-Making: An Overview.
- Mill, John S. [1859]/(1991) 'On Liberty', published in J. Gray (ed), *John Stuart Mill: On Liberty and Other Essays*. Oxford: Oxford University Press, pp. 1-128.
- Niesen, Peter (2004) 'Anti-extremism, Negative Republicanism, Civic Society: Three Paradigms for Banning Political Parties,' in Peer Zumbansen and Russell A. Miller (eds.), *Annual of German and European Law 2003*. Oxford: Berghahn, pp. 81-112.
- Oshana, Marina (1998) 'Personal Autonomy and Society', *Journal of Social Philosophy* 29.1: 81-102.
- Owen, Gareth S. *et al.* (2009) 'Mental Capacity And Judgmental Autonomy: An Interdisciplinary Challenge', *Inquiry* 52.1: 79-107.
- Owen, Gareth S., Cutting, John, and David, Anthony S. (2007), 'Are people with schizophrenia more logical than healthy volunteers?', *British Journal of Psychiatry* 191: 453-4.
- Pettit, Philip (1997) *Republicanism: A Theory of Freedom and Government*. New York: Oxford University Press.
- _____ (2001) *A Theory of Freedom: From the Psychology to the Politics of Agency*. New York: Oxford University Press.
- Rawls, John (1996) *Political Liberalism*. New York: Columbia University Press, [1993], Paperback Edition.
- Richardson, Geneva (2010) 'Mental Capacity at the margins: the interface between two acts', *Medical Law Review* 18: 56-77.
- Sadler, John Z. (2004), *Values and Psychiatric Diagnosis*. Oxford: Oxford University Press.
- _____ (ed.) (2002), *Descriptions and Prescriptions: Values, Mental Disorders, and the DSMs*. Baltimore: The Johns Hopkins University Press.

- Szasz, Thomas (1970), 'Involuntary Mental Hospitalization: A Crime Against Humanity', in his *Ideology and Insanity: Essays on the Psychiatric Dehumanization of Man*. London: Calder & Boyars.
- Sen, Amartya (2009) *The Idea of Justice*. London (etc.): Allen Lane.
- Tan, Jacinta O.A. *et al.* (2006) 'Competence to Make Treatment Decisions in Anorexia Nervosa: Thinking Processes and Values', *Philosophy, Psychiatry, & Psychology*13.4: 267-82.
- Tan, Jacinta O.A., Stewart, Anne, and Hope, Tony (2009), 'Decision-making as a Broader Concept', *Philosophy, Psychiatry, & Psychology*16.4: 341-4.
- Velleman, David (1999) 'A Right to Self-Termination?', *Ethics* 109.3: 606-628.
- Wakefield, Jerome C. (1992) 'The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values', *American Psychologist* 47.3: 373-88.
- Wilson, Derek (2007) *Out of the Storm: The Life and Legacy of Martin Luther*. London: Random House.
- Wolf, Susan (1989) 'Sanity and the Metaphysics of Responsibility', Christman (ed.) 1989: Ch, 9; reprinted from F. Schoeman (ed.), *Responsibility, Character, and the Emotions*, Cambridge: Cambridge University Press, 1987, Ch. 3.
- _____ (1990) *Freedom Within Reason*. Oxford: Oxford University Press.