

# *Capacity and Evaluative Judgment*

Binesh Hass

---

**Abstract**—This article advances two views on the role of evaluative judgment in clinical assessments of decision-making capacity. The first is that it is rationally impossible for such assessments to exclude judgments of the values a patient uses to motivate their decision-making. Predictably, and secondly, attempting to exclude such judgments sometimes yields outcomes that contain intractable dilemmas that harm patients. These arguments count against the prevailing model of assessment in common law countries—the four abilities model—which is often incorrectly advertised as being value-neutral in respect of patient decision-making both by its proponents and in statute.

## *1. Introduction*

When a patient with apparent cognitive impairment refuses treatment, healthcare providers are often required to assess the patient's 'decision-making capacity' (hereafter 'capacity'). If the result of the assessment is that the patient lacks capacity, then the decision to proceed with the treatment they refused will, subject to the relevant laws, sometimes fall to the healthcare providers or, in certain cases, the courts. A common worry in these cases is that the assessor may judge a patient to be incapable of providing informed consent merely because they disagree with the values a patient uses to motivate their reasoning. This worry has led to the rise of a class of value-neutral procedural models of capacity assessment which have become embedded in legislation and medical policy throughout the common law. The purpose of these models is twofold. First, they are meant to allow clinicians and the courts to avoid the thorny entanglements of evaluative judgment, that is to say, judgments about what is and is not of value in the lives of patients; and, second, they are supposed to yield better clinical outcomes than what would obtain under evaluative models of assessment. The objective of this article will be to show that these models fall short on both counts. They are neither able to help anyone rationally avoid the business of judging the values embedded in capacity assessment—which entails the necessity of deciding in favour of some values and against others—and nor do they yield better clinical outcomes than evaluative models.

I will start in section 2 with a couple of examples that will help determine what's at stake in making sense of the role of evaluative judgment in hard cases of capacity assessment. I follow in section 3 with a brief review of the key

premises of the prevailing value-neutral procedural models of capacity assessment, focussing in particular on one recent attempt to circumvent the evaluative worries I have described by grounding capacity assessment models on the idea of authenticity. This section will need to be brief and so general familiarity with the capacity assessment literature will be assumed. In sections 4 and 5, I will try to convince you that such models are not evaluatively neutral and, what's more, often result in intractable dilemmas that harm patients. I will conclude in section 6 with some remarks on how a straightforward evaluative model of capacity assessment which wears its values on its sleeves and is biased against serious prudential mistakes avoids such dilemmas.

## 2. Case Studies

The worry about the role of values arises in stark ways in particularly challenging capacity assessment cases,<sup>1</sup> such as the following:

- i. *Severe major depressive disorder.* A patient with treatable but lifelong severe major depressive disorder (MDD) refuses intervention options for early-stage cancer because they don't wish to live with depression and don't believe in the treatment options for it. The patient tells their doctor that the cancer will free them from the grips of depression.<sup>2</sup>
- ii. *Severe alcohol use disorder.* An epileptic patient with lifelong alcohol use disorder (AUD) is at risk of life-threatening alcoholic hepatitis. The patient is admitted to the hospital and subsequently experiences severe symptoms of alcohol withdrawal syndrome (AWS). As a result of not being able to access alcohol in the hospital, the patient states that they wish to forego treatment for alcoholic hepatitis and be discharged.

One might wonder why these examples don't constitute easy cases for capacity assessment in light of the fact that the patients' decisions emanate in large part from impaired cognitive or psychotic conditions.<sup>3</sup> That seemingly tempting

---

<sup>1</sup> At the time of writing, data on how often existing approaches are practically problematic for health practitioners is not readily available. With dementia in particular, however, roughly one in three cases have been self-reported by Swiss doctors as being 'challenging' on account of falling outside existing guidelines of DMC assessment; see C Poppe et al, 'Evaluation of decision-making capacity in patients with dementia: challenges and recommendations from a secondary analysis of qualitative interviews' (2020) 21 BMC Medical Ethics 1.

<sup>2</sup> Adapted, with important differences, from SYH Kim, 'The Place of Ability to Value in the Evaluation of Decision-Making Capacity' in DD Mosely and G Gala (eds), *Philosophy and Psychiatry: Problems, Intersections and New Perspectives* (Routledge 2016). In Kim's example, the patient's cancer is untreatable, whereas that is not the case here. I examine the implications of this change in treatability in §5.

<sup>3</sup> On MDD, see H Zuckerman et al, 'Recognition and Treatment of Cognitive Dysfunction in Major Depressive Disorder' (2018) 8 *Frontiers in Psychiatry* 1. On AUD, see R Rao and A Topiwala, 'Alcohol use disorders and the brain' (2020) 115 *Addiction* 1580.

route, however, is ruled out on account of *diagnostic neutrality*, which is the principle that no single diagnosis is sufficient for determining decision-making incapacity.<sup>4</sup> But if these patients seem unable to sufficiently value life over those lesser values, then why doesn't that constitute sufficient grounds for determining incapacity? The answer to that, in turn, centres on the constraint of *value neutrality*, which, like the diagnostic constraint, rules out a verdict on the basis of evaluative judgments alone and thus acts as a stop against undue paternalism. These observations take us nicely to the starting point of the value-neutral procedural models whose chief ambition, as I say, is to allow clinicians to avoid the entanglement of evaluative conflicts and, by so doing, prevent suggestions of undue paternalism.

### 3. *Abilities and Values*

The most influential model of capacity assessment throughout the common law is Grisso and Appelbaum's four abilities model, which centres on an individual's abilities (i) to understand the information that is provided to them, (ii) to appreciate the facts that apply to them, (iii) to reason with such facts, and (iv) to communicate in some way their preference about the options before them.<sup>5</sup> Together, these abilities are jointly necessary to establish that an individual has capacity. Part of what motivates the model is the view that the way to assess capacity ought to be process- rather than outcome-oriented. That is to say, the assessor should be concerned with how an individual arrives at a decision, if they do so at all, rather than with the content or consequences of the decision itself. The assumption has been that this is the best way to safeguard a person's autonomy in contexts where there is a real risk of infringement by medical or legal authorities. Yet despite the many insights and important work this model has enabled, a growing body of academics and clinicians are pointing out the various ways in which it fails to capture a range of cases where patients clearly lack capacity. These lapses have been particularly acute when cases have directly pertained to the values that patients use to motivate their decision-making when they refuse treatment.<sup>6</sup> The need to address such lapses has in turn resulted in various efforts to upgrade the four abilities model with more sensitive evaluative

---

<sup>4</sup> J Hawkins, 'Affect, Value and Problems Assessing Decision-Making Capacity' (St Cross Special Ethics Seminar, Oxford, 19 November 2020).

<sup>5</sup> PS Appelbaum and T Grisso, 'The MacArthur Treatment Competence Study' (1995) 19 *Law and Human Behavior* 105.

<sup>6</sup> JOA Tan et al., 'Competence to make treatment decisions in anorexia nervosa: Thinking processes and values' (2006) 13 *Philosophy, Psychiatry, & Psychology* 267; NF Banner and G Szmukler, 'Radical Interpretation' and the assessment of decision-making capacity (2013) 30 *Journal of Applied Philosophy* 379; G Szmukler, 'Anorexia Nervosa, Lack of "Coherence" with Deeply Held Beliefs and Values, and Involuntary Treatment' (2021) 28 *Philosophy, Psychiatry, & Psychology* 151; JH Radden, 'Food Refusal, Anorexia and Soft Paternalism: What's at Stake?' (2021) 28 *Philosophy, Psychiatry, & Psychology* 141.

apparatus while at the same time attempting to retain the model's process-oriented and value-neutral ambitions.

One influential suggestion along these lines has been Scott Kim's thesis that a patient's ability to value—but not the cogency of the content of those values—ought to constitute one further jointly necessary condition for the determination of capacity.<sup>7</sup> This, in essence, is a test for *evaluative incapacity*, the determination of which must satisfy two conditions. First, the patient's decision-making in contexts where informed consent is necessary for treatment must be reasonably determined to stem from 'some type of pathology, psychiatric or otherwise, or malfunctioning of brain processes'; and second, the patient must be subjectively inconsistent on significant matters of value, which is meant to indicate that something has gone seriously awry in the underlying ability to value.<sup>8</sup> The second condition is elaborated through the idea of authenticity, which works like a counterfactual presumption. For example, in the case of the patient with MDD, the assessor is permitted to presume, in light of the test's first condition, that the patient would not value death over life if it were not for the fact that they are severely depressed. The underlying psychopathology renders the values that motivate the patient's decision to forego treatment for early-stage cancer inauthentic insofar as they stem from the depression. And so, the evaluative judgment that produces that decision is best seen, in Kim's words, 'as an enemy within'.<sup>9</sup> It is 'inauthentic' on this view because it is not what the patient would want if they were not affected in such a severe way by MDD. When the inauthenticity is that bad, it is tantamount to incapacity according to Kim's model.

#### 4. *Values and Reasons*

By directing attention to the consistency of a patient's values across time and away from the actual content of the value, the hope is that capacity assessments are less likely to fall afoul of the prevailing ideal that medical authorities should allow people to live their lives according to their own beliefs. The immediate payoff of the model, according to its proponents, is that it allows assessors to avoid the messy business of making sense of, and judging, patients' values. But the easy allure of that immediate payoff conceals deeper problems. Let me point out three.

First, procedural models do not exist in a value-neutral vacuum. They are themselves justified, at base, by some conception of what would be valuable to promote by setting up the procedures in the first place. All positive rules, of which procedures are one kind, are like this. The evaluative ideas deployed in the premises of the justifications for such rules moreover are transitive. If a set

---

<sup>7</sup> See S Kim, *Evaluation of Capacity to Consent to Treatment and Research* (OUP 2010) and his 'The Place of Ability to Value in the Evaluation of Decision-Making Capacity'.

<sup>8</sup> *ibid* 195.

<sup>9</sup> *ibid* 196.

of procedures is established to promote some particular value, then each discrete procedure from that set will promote that value in some way. For otherwise that procedure's inclusion in that set would be rationally pointless. The result of that piece of reasoning is that there is no durable boundary between procedural and substantive rules, since all rules turn out to be at least minimally substantive insofar as they promote their initial justifications. Now, some legal philosophers have tried to exclude the transitivity between rules and their justifications in the hopes of grounding rules as reasons for actions themselves (the reflexivity thesis) rather than as merely paraphrased statements of their justifications (the paraphrastic thesis).<sup>10</sup> Proponents of the four abilities model might be tempted to assume the reflexivity thesis as support for the value-neutrality ambitions of their model. This would be an error, however. The reflexivity thesis is concerned with severing the connection between a rule and its evaluative justification for the purpose of establishing the rule as a reason for action in its own right. But that thesis does not entail a denial of the evaluative implications of the rule as a reason for action, for that is a very different kind of implication than merely bracketing transitivity in the context of practical reasoning. My own view, which I have defended at length elsewhere,<sup>11</sup> is that the reflexivity thesis rests on a logical mistake and that legal rules taken by themselves are bereft of rational force—but even granting the reflexivity thesis in *arguendo* will not, as I say, yield the value-neutrality desired by some advocates of the four abilities model. For the key claim of those who defend the model is that its capacity assessment process is neutral in respect of the values embedded in the patient's decision-making, *even if* the assessment process is itself at base not evaluatively neutral. And so, according to its advocates, the model achieves the desired neutrality by excluding the values embedded in the patient's decision-making from the assessment process. Yet the neutrality is illusory. For any set of values *A* that is embedded in the patient's decision-making will need to stand in some kind of relation to the set of values *B* that justify the process of assessing that decision-making in a particular way. For practical purposes, it is possible for *A* and *B* to be evaluatively neutral in respect of each other only if they have nothing to do with another, i.e., there is no interaction between the sets of values. But this obviously cannot be the case where *B* justifies a particular process of capacity assessment that explicitly stipulates that *A* is not to be taken into consideration. Exclusion from this kind of consideration is evaluative in at least two senses: first, that *B* is regulating how *A* is to be treated rather than, for example, the other way around, reveals a certain ethical priority; and second, the capacity assessment that results from excluding *A* may have significant ethical consequences. So, clearly, nowhere in this story is *A* treated in a way that even

---

<sup>10</sup> J Raz, 'Reasoning With Rules' in J Raz, *Between Authority and Interpretation* (OUP 2011) and J Rawls, 'Two Concepts of Rules' (1955) 64 *The Philosophical Review* 3.

<sup>11</sup> B Hass, 'The Opacity of Rules' (2021) 41 *Oxford Journal of Legal Studies* 407.

resembles value-neutrality.<sup>12</sup> This observation takes us to the second problem with seeing value-neutrality where there is none.

In models where authenticity or consistency in one's 'internal rationality' is necessary for a determination of capacity,<sup>13</sup> clinicians need to ascertain whether the patient's refusal of a particular treatment option reflects a value that is inconsistent with the patient's other values both more generally and across time. That would be the clinical test, for Kim in particular, of the patient's ability to value in the first place. A pair of presuppositions packaged with this test will strike many observers as bizarre and problematic, which is that consistency across time is a good way to gauge someone's ability for evaluative judgment and, by extension, that inconsistency will indicate that something has gone awry. What this rules out as eligible evaluative data points in a patient's history are moments of clarity brought about by, say, near-death or life-altering experiences such as one might expect to find on a routine basis in clinical settings. For one would have very good reasons to ditch a hitherto held set of supposed values in contexts where one's very existence is threatened by those values. To then have that clarity called into question by a model that privileges consistency in one's judgment and which then deploys a finding of inconsistency as grounds for determining incapacity would yield a result that most observers would find unacceptable. It is unacceptable because it would entail a clear and serious breach of a patient's right to autonomy, the very thing the model sets out to safeguard, and, what is more, it elevates consistency in evaluative judgment as a value itself—but obscured by descriptions of value-neutrality and hidden, as it were, in the procedural model's unstated premises rather than out in the open where it can be debated by the patient, their family, and other concerned parties.

The theoretical problem of hidden evaluative premises belies practical quandaries. For one further issue with consistency is that it won't help anyone as a metric for capacity in contexts where a patient is known to consistently make serious prudential mistakes across time (hereafter 'prudential mistakes'). These are, as Jennifer Hawkins puts it, mistakes that are (a) virtually irreversible and (b) leave the patient seriously worse off than they would have been had they (c) chosen an alternative course of action that was (d) relatively easy to act

---

<sup>12</sup> This is not the only way to expose the evaluative commitments of putatively procedural assessment models. Natalie Banner and George Szmukler, for example, have argued that assessors necessarily assume some form of Donald Davidson's principle of charity, which is a normative attitude of interpretation that assumes that a speaker is reasonably consistent in their beliefs and that such beliefs correspond to what is true. See NF Banner and G Szmukler, "Radical Interpretation" and the Assessment of Decision-Making Capacity' (2013) 30 *Journal of Applied Philosophy* 379, 383.

<sup>13</sup> L Charland, 'Mental competence and value: The problem of normativity in the assessment of decision-making capacity' (2001) 8 *Psychiatry, Psychology and Law* 135, 136 on 'internal rationality'.

upon.<sup>14</sup> Cases in which patients fit that bill happen all the time and yet that kind of unenviable consistency would, on the authenticity model, tag a standout evaluative judgment that gets things right at some crucial clinical juncture as being indicative of incapacity rather than of a moment of clarity to be seized upon. No doubt that bizarre outcome is likely to strike many as unacceptable both as a theoretical matter and, as I will suggest in the following section, for the fact that it results in worse practical outcomes for the patient than what would be the case under an evaluative model of capacity assessment that is biased against prudential mistakes.

### 5. *Values and Outcomes*

Consider the patient with treatable but lifelong severe MDD who has an early-stage cancer which also has a good chance of being successfully treated. It is not uncommon for such patients to regard their depression as essential to their identity.<sup>15</sup> Yet it is easy to see how an authenticity model would struggle to parse the patient's symptomatic expressions of MDD, such as an unreasonable outlook of despair about treatment options, from what the patient could reasonably be assumed to believe about the prognosis of their cancer in the absence of their depression. Because the patient has had MDD for so long and associates it so closely with their identity, a model of authenticity that is looking for inconsistency in the patient's evaluative judgements as a foothold for assessing incapacity is going to be a non-starter. There will be no inconsistency in the patient's evaluative judgements because the sense of despair that is likely causing the patient to make the prudential mistake of refusing treatment will be consistent with the patient's more general worldview. From an outcome perspective concerned with patient wellbeing, the story gets worse because there is no available route to providing treatment through a determination of incapacity. Without treatment, and other things being equal, the patient will very likely die of cancer.

You might wonder about the role that patient–doctor discussions could take in cases such as these. Suppose that the discussions are extensive and clarify that the patient's beliefs in respect of their decision to refuse treatment reflect values that, upon further consideration, are recognised by the patient themselves not to comport with their values more generally. Let's say that those discussions make the patient recognise that their true values don't count in favour of the decision to refuse treatment, leading them to change their mind in that regard. So, you might say in such a case, if it were not for the ability-to-value assessment and the extensive discussion it occasioned, we may have missed out on getting

---

<sup>14</sup> Hawkins, 'Affect, Value and Problems Assessing Decision-Making Capacity'. See also J Hawkins, 'Why even a liberal can justify limited paternalistic intervention in anorexia nervosa' (2021) 28 *Philosophy, Psychiatry, & Psychology* 155, 156 f.

<sup>15</sup> T Cruwys and S Gunaseelan, "'Depression is who I am": Mental illness identity, stigma and wellbeing' (2015) 189 *Journal of Affective Disorders* 36.

the patient to see that the treatment is desirable even from their own evaluative point of view. That is no doubt a positive outcome. But notice that it isn't the assessment that is doing the work here in getting us to that outcome. It is the discussion. And that discussion should be occurring, preferably with an ethicist, even in the absence of any test of the ability to value. There will, however, be other cases where these discussions reveal that the patient's decision to refuse treatment, in fact, reflects a considered judgment that comports with their evaluative worldview. And if, in addition to that evaluative consistency, the patient's MDD is both severe and lifelong, then on what grounds could anyone say that the patient, in refusing treatment, is making prudential mistake even if their decision emanates from the MDD? Just to be clear, whichever way one treats the patient's decision, one is already engaging in some rather heavy duty evaluative judgment. In light of the life-and-death stakes involved for this patient, one will either need to adjudge that the patient is *not* making a prudential mistake in refusing cancer treatment against a backdrop of their MDD *or* that they are making such a mistake.<sup>16</sup> There is no value-free third way if the patient's decision is to be regarded as a reason for action. The question, then, is not whether but which evaluative framework should be brought to bear on that decision by healthcare providers and the courts in deciding which values should prevail when they conflict. The framework of prudential mistakes is one such evaluative framework that prioritises preserving life as the default option. It is also a default that is consonant with what is likely the most important ethical commitment of contemporary medical practice. That evaluative commitment—to preserve life—provides *prima facie* grounds for regarding the patient's decision as a prudential mistake.

There are, however, other considerations that count in favour of an evaluative capacity assessment model. Take the case of the patient with severe and lifelong AUD who is at risk of life-threatening alcoholic hepatitis. Their admission to the hospital produces severe AWS symptoms, including episodes of delirium tremens and a persistent urge to re-establish access to alcohol. As a result of that persistent urge, and upon recovering from the delirium, the patient indicates that they wish to be discharged and hence forego treatment for alcoholic hepatitis. Suppose that the severity of these symptoms, along with other concerns, prompts a capacity assessment even in the absence of alcohol-induced brain degenerative disease, such as Korsakoff's Syndrome. Now consider two possible outcomes: first, discharging the patient without treating the alcoholic hepatitis is likely to result in the patient's death as a result of liver failure;<sup>17</sup> and second, discharging the patient without treating the alcoholic hepatitis won't result in the patient's death as a result of liver failure but it will,

---

<sup>16</sup> For a discussion of why selecting between alternatives is always evaluative, see Hass, 'The Opaqueness of Rules', 411–12.

<sup>17</sup> TS Sehrawat, M Liu, and VH Shah, 'The knowns and unknowns of treatment for alcoholic hepatitis' (2020) 5 *The Lancet Gastroenterology & Hepatology* 494.



other things being equal, almost certainly intensify the underlying alcohol dependence and, particularly in light of the patient's history of epilepsy, lead to further near-term life-threatening conditions. It is clear that neither possibility is desirable from an outcome-based point of view that is concerned with the patient's wellbeing. Yet an autonomy-maximalist might think that these outcomes, at the very least, respect the right of the patient to decide for themselves how they wish to live their lives, however ill-advised that might be. This line of thought is both common and regrettably question-begging. The overlapping theoretical consensus in the literature on autonomy is that, at the very least, autonomy is pegged to the capacity for reasoning—the more one has of the latter, the more one partakes in the former.<sup>18</sup> This degree-like nature of autonomy is an important clue to understanding its misuse in settings where the status of consent is relevant. A patient reeling from AWS, for example, will be quite a long way from a state of mind that would enable them to engage in a degree of reasoning relevant to qualifying them as autonomous.<sup>19</sup> Let me call this idea *negative autonomy* and put it this way: if a patient is not sufficiently autonomous *from* symptoms that are debilitating for reasoning, then they are not sufficiently autonomous *for* decisions that can be reasonably inferred to have been produced by those symptoms.<sup>20</sup> And if they are not already autonomous in that specific respect, it is question-begging to use autonomy as a premise for respecting the patient's wish to be discharged and hence forego treatment.<sup>21</sup>

One might wonder how a value-neutral authenticity model could get a foothold in the kind of case I just described. If, on the one hand, (a) a patient's AWS is debilitating their reasoning and, by extension, their autonomy, then, fortunately for the patient, that might suggest that a determination of incapacity is a real possibility on those grounds. But it isn't—for, on the other hand, (b) the patient's AUD is lifelong, and possibly exacerbated by other reason-debilitating comorbidities, meaning that (i) the patient's negative autonomy has long been in doubt and, as a result, (ii) their refusal of treatment is likely to be consistent with similar decisions and refusals in the past. In particular, (ii) makes the reasoning that positively values re-establishing access to alcohol and negatively values states in which that isn't possible, such as remaining in the hospital, as 'authentic' for the patient in question in light of their history. And so, the answer to the puzzle of how a value-neutral authenticity model might

---

<sup>18</sup> G Owen Schaefer, G Kahane, and J Savulescu, 'Autonomy and Enhancement' (2014) 7 *Neuroethics* 123, 126.

<sup>19</sup> Rao and Topiwala, 'Alcohol use disorders and the brain' op. cit.

<sup>20</sup> The language of desire is adopted merely because it is idiomatic, but it is clear that someone in the grips of a severe dependency often doesn't desire the thing to which they are dependent, which is a fact that bears out the unfreedom of pathology. Cf J Savulescu, 'Rational Desires and the Limitation of Life-Sustaining Treatment' (1994) 8 *Bioethics* 191.

<sup>21</sup> The Australian Law Reform Commission has noted that, in instances where it is not possible to ascertain what the patient actually wants even absent disruptive symptoms, considerations of the patient's human rights

resolve such cases is that it can't. This result is undesirable given the outcomes it entails for the patient.

## 6. *Conclusion*

Part of what makes 'value-neutral' procedural models inadequate for dealing with the kinds of cases under discussion is that they don't provide a route to evaluatively judge the relevance of prudential mistakes when such mistakes form an essential part of a patient's history. As I alluded earlier in this article, a model that does make space for that kind of evaluative judgment is, unsurprisingly, an essentially and straightforwardly evaluative model of capacity assessment. It is evaluative because it will involve judgment about what values should prevail when they conflict (e.g., the value of being discharged and regaining access to alcohol versus the value of not being discharged and receiving treatment for a life-threatening condition). And the model is *essentially* evaluative because the evaluative judgment it entails is a necessary condition of the capacity assessment. It is necessary because one would otherwise end up at the common dilemma captured by (a) and (b) in my previous section. Avoiding that dilemma is useful in itself but perhaps an equally useful feature of a model that allows for the evaluative judgment of prudential mistakes is that is very obviously evaluative. It thus clears away the confused ambitions of value-neutrality and, by so doing, enables space for greater sophistication in ethical debates in clinical and legal settings about capacity assessments.