Ethical Theory and Moral Practice (2025) 28:139–153
https://doi.org/10.1007/s10677-024-10471-y

Capacity, Autonomy, and Risk: Reflecting on Asymmetries in Capacity to Consent and Capacity to Refuse

Jonathan Pugh[1]

Accepted: 9 October 2024 / Published online: 28 October 2024

Abstract

There has been renewed interest in whether we should understand standards of decision-making capacity (DMC) to be risk-relative. Critics of risk-relative standards often highlight a puzzling asymmetry that they imply; a patient may have the requisite DMC to consent to a treatment that is in their best interests, whilst lacking the requisite DMC to refuse that same treatment, given the much higher risk that this would entail. Whilst some have argued that this asymmetry suggests that risk-relative standards are nonsensical, in this paper I defend a 'quality of evidence' view of such standards. I begin by outlining DMC's purported gate-keeping role in medical ethics, and identifying three key normative claims that undergird this role. I then explain how two competing theories of risk-relative standards are incompatible with at least one of these claims. Drawing on Douglas' distinction between standards of 'true capacity' and standards invoked in the 'test' for capacity, I then outline my 'quality of evidence' view. I explain how the view is compatible with the aforementioned normative claims, and outline the nature of the asymmetry it implies. I conclude by responding to the objection that there is no meaningful distinction between 'true capacity' and the 'test' for capacity.

Keywords Decision-making capacity · Autonomy · Risk · Consent · Medical ethics

The concept of decision-making capacity (DMC) is often understood to play a 'gate-keeping' role in medical ethics (Beauchamp and Childress 2001, pp. 69–73; Buchanan and Brock 1989). This role can be understood to have a positive and a negative aspect. The positive aspect is that if a patient is found to have DMC to consent to (or refuse) a medical treatment, then they are often understood to have the right to make their own decision about whether to undergo that treatment, even if their decision might be contrary to their best interests. Consider, for instance, cases in which physicians might respect a Jehovah's Witness' refusal of a life-saving blood transfusion (Bock 2012). The negative aspect of DMC's gate-keeping

---

✉ Jonathan Pugh
jonathan.pugh@uehiro.ox.ac.uk

[1] Uehiro Oxford Institute, University of Oxford, Suite 8, Littlegate House, St Ebbes Street,
Oxford OX1 1PT, UK

role implies that decisions made by a patient who *lacks* DMC may not always be respected. Instead, treatment decisions for such patients are often understood to be governed by considerations of beneficence (Birchley 2021; Buchanan and Brock 1989).

There is scope to question whether DMC should have this gate-keeping role. Indeed, it is challenged by the Convention on the Rights of Persons with Disabilities, (United Nations 2006) and some authors have recently argued that we should abandon the concept in clinical practice (Fogal and Schwan forthcoming). Whilst recognising the importance of these issues, in the interests of brevity, I must set them aside. Instead, I shall consider the narrower question of whether we should understand standards of DMC to be risk-relative, given its gate-keeping role; should riskier decisions be understood to require a higher threshold of DMC than decisions with less significant consequences?

There has recently been renewed interest in this question. Although the risk-relative view has been invoked in influential legal judgments, and received considerable philosophical support, (Lawlor 2016; Manson 2015; Wilks 1999) others have recently argued that it is conceptually misguided to conflate risk with DMC, joining previous theorists in claiming that risk-relative standards are nonsensical (Buller 2001; Cale 1999; Culver and Gert 1990; Harris 2003; Pickering et al. 2022; Wicclair 1991). Discussions of risk-relative standards typically focus on a puzzling asymmetry that the view implies (Lawlor 2016; Manson 2015) – on such standards, a patient may have the requisite DMC to consent to a treatment that is in their best interests, whilst lacking the requisite DMC to refuse that same treatment, given the much higher risk that this would entail. To illustrate the asymmetry, suppose Joe requires a straightforward, safe, but potentially life-saving surgery (say to remove relatively accessible cancerous tissue, following early diagnosis). In this case, the point of the asymmetry implied by risk-relative standards is that Joe could be found to have DMC to consent to this surgery, and yet be found to lack DMC to refuse it due to the different risks associated with each decision.

In this paper, I shall defend a ‘quality of evidence’ view of risk-relative standards of DMC – on the approach I defend, we may justifiably require a higher quality evidential basis for riskier decisions. I shall argue that this view can justify a version of the asymmetry highlighted above, and that, contrary to competing justifications of risk-relative standards, my approach is compatible with key normative claims that undergird DMC’s gate-keeping role. I shall begin by outlining a justification of this role in more detail, before identifying the normative claims that it reflects. I shall then outline prominent ways in which risk-relative standards have been defended and show how existing attempts are incompatible with the aforementioned normative claims. In the final section, I will defend my ‘quality of evidence’ view.

1 An Autonomy-Based Justification of DMC’s Gate-Keeping Role

To begin, a note on terminology; I shall parse the arguments about risk-relative standards in terms of ‘capacity’. These arguments are instead sometimes parsed in terms of ‘competence’. Whilst these two terms are often used interchangeably, they can also denote separate legal and medical conceptions of an individual’s decision-making abilities (Pugh 2020, p. 184). However, the terms are used for these purposes inconsistently across different jurisdictions. To avoid these complications, I will simply use the term ‘DMC’ to refer to the property that is understood (in either a medical or legal context) to be necessary for a patient to make a medical treatment decision that warrants respect. I shall now explain how considerations of autonomy can provide plausible grounds for why DMC might be necessary for one's treatment decisions to warrant respect.

As detailed above, patients with DMC are afforded considerable authority to make their own medical treatment decisions, even when this is against their best interests. We can imagine an alternative in which decisions about medical treatment were instead governed by considerations of beneficence alone. Mill famously rejected the paternalism that would be inherent in this alternative approach in his harm principle, according to which the individual's "own good, either physical or moral, is not a sufficient warrant" to justify an interference with that individual's liberty (Mill 2003, p. 18). In a related vein, we may say that one reason why this sort of paternalistic approach to medical decision-making is largely rejected in contemporary medical ethics is because it overlooks the salient value attributed to personal autonomy.

Autonomy is a highly contested concept, and I cannot outline a full theory of it here. We may simply acknowledge that autonomy is often understood in a *procedural* sense in contemporary ethics. On such an approach, autonomous decisions must be made in accordance with a certain kind of decision-making procedure; crucially, procedurally autonomous choices need not align with what is in an individual's best interests.

The precise nature of the procedure that autonomous decision-making involves will depend on the details of one's chosen theory. However, one widely accepted procedural condition is that the individual's decision must be grounded by a sufficient understanding of information that is material to their decision (Beauchamp and Childress 2001). After all, in order to make a decision that reflects our values, in the manner that connotes the self-governance that the concept of autonomy aims to capture, we plausibly need to understand what our options are like, and how they relate to our values (Savulescu and Momeyer 1997). Indeed, it is widely accepted that autonomy, understood in this broadly procedural sense, provides a moral foundation for informed consent. Informed consent, on one plausible understanding, can be construed as giving an 'autonomous authorisation' of a certain treatment (Faden & Beauchamp, 1986).

This discussion raises two questions; first, what is the relationship between DMC and procedural autonomy? Second, why are considerations of autonomy understood to outweigh considerations of beneficence in a manner that justifies DMC's gate-keeping role? I shall consider each in turn.

As detailed above, procedural theories of autonomy set out various conditions that must be met if a decision is to qualify as autonomous. Drawing on Friedman's terminology, we may say that these theories outline *constitutive* conditions of autonomous decision-making (Friedman 2003, p. 4); so, in light of the brief analysis above, we might plausibly say that autonomous decision-making is partly constituted by deciding on the basis of a sufficient understanding of material information.

To make a decision that meets the constitutive conditions of autonomy, agents will need to have certain abilities. Again, to use the example of understanding, we may say that an individual cannot make an autonomous decision if they lack the ability to understand information that is material to that decision. Accordingly, and again drawing on Friedman's terminology, we can understand the concept of DMC to pertain to the abilities that are *causally* necessary for making a decision in accordance with a procedural account of autonomy

(Friedman 2003, p. 4). On such an approach, DMC is to be construed in a decision-specific manner, and the fact that a person makes a decision that is contrary to their best interest does not alone entail that they lack DMC. Instead, a patient will only lack DMC if they lack the abilities that are causally necessary for making a particular decision in a certain way.

I have not outlined precisely what abilities DMC requires. This is because, on the understanding I am outlining here, the requirements of DMC will ultimately depend on one's theory of the constitutive conditions of autonomy. Due to the contentious nature of autonomy, I do not wish to commit myself to a particular view about these conditions. For the purposes of this paper, I do not need to; what matters for my purposes is that there is an inextricable relationship between how we should understand DMC, and how we should understand what constitutes an autonomous decision. Moreover, we may note that although the abilities that might be deemed necessary for autonomous decision-making will often admit of degree, DMC is best construed as a range property; although we may assign it based on the individual's possession of various scalar abilities, DMC is itself a binary property that is possessed once an individual passes a certain threshold of the relevant abilities (McMahan 2002, p. 250).

As I shall explore in the next section, one of the key questions for a theory of DMC is where to set this threshold. However, it is important to note that there is a crucial difference between having DMC with respect to a particular decision and being *known* to have it. Indeed, third parties face significant obstacles to ascertaining whether another person has DMC with respect to a particular decision. This is not only due to disagreements about the nature of autonomy, but also because many of the abilities relevant to DMC cannot always be straightforwardly assessed (Douglas 2022). As such, we might usefully draw a distinction between what Douglas calls (i) standards of 'true' DMC, and (ii) standards we invoke in the 'tests' we employ in assessing DMC (Douglas 2022). The former relates to the abilities that suffice for a person to be autonomous with respect to a decision. In contrast, the test for capacity relates to the ". . level of capacity a person must evidence if we are to treat her decision *as if it were* autonomous" (Douglas 2022, p. 800). I shall return to this distinction below, as it plays an important role in both Douglas' defence of a risk-relative standard, and my own. Henceforth, though, when I refer to DMC, I shall mean to refer to 'true DMC' unless stated otherwise.

This exploration of the relationship between DMC and autonomy can provide us with a normative explanation of the *negative* aspect of DMC's gate-keeping role. When a patient lacks DMC, they lack abilities that are necessary for them to make an autonomous decision to refuse (or consent to) treatment; in the absence of any other plausible basis for presuming consent, there are no countervailing autonomy-based considerations that can outweigh beneficence-based arguments in favour of providing (or withholding) medical treatment. Accordingly, the negative aspect of DMC's gate-keeping role can be construed as a form of soft paternalism; the idea that one can permissibly interfere with an agent's non-autonomous choices to protect their interests (Feinberg 1984, pp. 12–13).

Conversely, the positive aspect of DMC's gate-keeping role enshrines a rejection of hard paternalism: the view that one may permissibly interfere with an agent's autonomous choices to protect their interests (Feinberg 1984, pp. 12–13). We still need an explanation of this aspect; why should reasons to respect autonomous choice override reasons of beneficence? Such a position can be defended by appealing to both the moral and prudential value of autonomy; again, I need not commit myself to any particular view here. However, it will be useful to briefly highlight one view that fails to offer a convincing account of the salience of autonomy's value, namely what Ronald Dworkin terms the 'evidentiary view' (Dworkin 1993, pp. 119–123).

According to this view, our reasons to respect autonomy are grounded by the fact that the individual herself is typically the best judge of what is in her own interests. The evidentiary view does capture an important truth here; however, as Dworkin recognises, it does not capture the whole truth. To illustrate, Dworkin describes an individual who smokes despite knowing that smoking is not in their best interests (Dworkin 1993, p. 180) If we believe that we ought to respect the individual's choice to smoke, it cannot be because their choice provides the strongest evidence of what is in their best interests. The point of this argument is that the salient value placed on respecting autonomy is not reducible to the fact that autonomous choice serves as a reliable guide to what is in our best interests; autonomy must instead have some deeper value.

Fortunately, there are several alternatives to the evidentiary view that can account for this deeper value. Dworkin himself favours what he calls the integrity view; here, autonomy is understood to have value because it is essential for living a life structured by our own values (Dworkin 1993, pp. 120–123). Of course, there are various other ways of justifying the claim that considerations of autonomy should override those of beneficence, including those sourced from Kantian philosophy that place greater emphasis on autonomy's moral value (Cholbi 2017). Whilst I do not need to commit myself to a particular view, I do want to suggest that the evidentiary view provides an unsatisfactory account. Instead, it appears to be in tension with the positive aspect of DMC's gate-keeping role, in so far as individuals with DMC are afforded the authority to make decisions that are contrary to their best interests.

We can now identify three interrelated claims that undergird DMC's gate-keeping role in medical ethics on the autonomy-based approach outlined here. Although I have not defended them here, I believe that we should accept these claims. In any case though, a plausible theory of risk-relative standards of capacity must either be compatible with these claims, or provide a revisionary account of why they should be rejected:[1]

> *The Procedural Autonomy-Capacity Claim*: DMC pertains to the abilities that are causally necessary to making a procedurally autonomous decision, where such decisions can be contrary to the individual's best interests.
>
> *The Anti-Evidentiary Claim*: Our reasons to respect autonomy are not reducible to the fact that autonomous choice is a reliable guide to what is in the individual's best interests.
>
> *The Anti-Hard Paternalist Claim*: Our moral reasons to avoid failures to respect an individual's autonomous choice have priority over our reasons to avoid failures to act in that individual's interests.

---

[1] A reviewer of this manuscript enquired about the implications of risk-relative standards and their problems for relational accounts of autonomy. Whilst this is an important question given the interest in such theories, it is difficult to offer a response that is both comprehensive and concise, as 'relational autonomy' covers such a broad church of theories. However, there is no inherent reason why relational theories per se could not endorse a risk-relative standard of DMC, and encounter the problems outlined here. However, certain relational theories could be immune to these problems if they offer a revisionary explanation for rejecting any of the three claims outlined here (such as certain substantive relational theories).

In what follows I shall seek to show how the risk-relative approach to DMC can be compatible with these claims.

2 Rejecting the 'True Capacity Threshold View' and the 'Cost of Error View'

Recall that risk-relative standards of DMC imply a puzzling asymmetry: a patient may be found to have DMC to consent to a given medical treatment, and yet be found to lack DMC to refuse that same treatment. Following the above analysis, we can immediately identify one reason why this appears problematic. As Manson puts the point:

> In order to make a rational decision about a course of action, we need to be able to comprehend both the course of action (and its implications) and the implications of inaction. (Manson 2015, p. 68)

Accordingly, the asymmetry appears to be at odds with the 'procedural-autonomy capacity claim' identified above. If DMC pertains to the abilities that are necessary to make a procedurally autonomous decision, then it is unclear why there should be an asymmetry between the DMC required to consent to and refuse the same treatment (such as Joe's surgery in my initial example). Making either decision autonomously will plausibly require the very same degree of the ability to understand information, because the exact same information is material to both choices.

However, this concern need not raise problems for more nuanced interpretations of the asymmetry. Lawlor claims that the concern only undermines the view that an individual could *possess* some degree of the abilities required to consent to treatment autonomously, and yet possess a lesser degree of the abilities required to autonomously refuse treatment (Lawlor 2016, p. 749). Crucially though, Lawlor claims that this interpretation is not the one that defenders of risk-relative approaches mean to defend; rather, the interpretation they seek to defend pertains to the different thresholds of DMC that individuals must pass in order to qualify as having the DMC to make different kinds of decision. So, the idea here is not that the person possesses different degrees of the abilities relevant to DMC in cases of consenting to and refusing treatment. Instead, the claim is that the degree of the relevant abilities that an individual possesses might confer true capacity to consent to treatment; yet those same abilities may not similarly confer true capacity to refuse, if this requires that one passes a higher threshold. Call this the 'true capacity threshold view' of the asymmetry – the true capacity required for a decision to refuse treatment may require a greater degree of the relevant abilities than the true capacity required for a decision to consent to a treatment.

This view is not vulnerable to Manson's concern, but why should we accept it? Why suppose that we should adopt a higher threshold of true capacity for riskier decisions? In defence, Lawlor appeals to two thought experiments where analogous asymmetries seem appealing. The first is drawn from Wilks:

> I suggest that an acrobat of middling reliability might be foolhardy to walk the line without a safety net strung below, while there might be nothing untoward in their attempting it with net in place; and I see it as entirely in accordance with correct usage

> to describe this as a case where the acrobat is competent to walk the line in the second instance but not in the first. (Wilks 1999, p. 156)

The second is drawn from a comedy routine in which a subject of an inquisition is given the choice between 'cake' or 'death'. Here, Lawlor suggests that someone could have the requisite abilities for having the DMC to choose cake, and yet these same abilities would not be sufficient for them to pass a higher threshold of DMC to choose death (Lawlor 2016, pp. 751–752).

Yet, notwithstanding the intuitive pull of these examples, there are various problems with the true capacity threshold view. In addition to questioning Lawlor's reliance on intuitions, Pickering et al. note that the thought experiments beg an important question by tying standards of DMC to safety (Pickering et al. 2022, pp. 894–895). I am sympathetic to this concern; it certainly seems as if the examples are in tension with the procedural-autonomy claim I have outlined above. Nonetheless, I suspect that Lawlor might respond to this line of objection by claiming that, rather than begging the question, his examples serve to reverse the burden of proof for what does and does not constitute DMC. To illustrate, the pull of the acrobat example is partly grounded in the fact that our concept of what it is to be a competent tightrope walker (or to have the capacity to walk a tightrope) involves the ability to perform the feat *without causing significant harms to ourselves*. So, Lawlor might quite plausibly ask his opponent to show why our concept of DMC should be different.

However, this is a burden of proof that Lawlor's opponent can meet (and without making any claims about the validity of thought experiments in moral epistemology). The acrobat example fails as an analogy for our understanding of DMC because procedural conceptions of autonomy *deny* that autonomous decisions cannot lead to significant harm. Accordingly, one can have the ability to decide autonomously without having the ability to make a 'safe' decision on these views, even if one cannot similarly be said to have capacity to walk a tightrope whilst lacking the ability to do so safely.

As such, the true capacity threshold view is in tension with the procedural-autonomy claim; moreover, other seemingly promising lines of defence for the claim that DMC should be tied to considerations of safety are incompatible with the two other key normative claims I identified in the previous section. To illustrate, one might argue that DMC should be tied to considerations of safety, if our reasons to respect autonomy were reducible to the fact that autonomous decision-making is a reliable guide to what is in a person's best interests. If that were so, then supplementing our understanding of DMC with considerations of safety might restrict the treatment decisions that are respected to those in which autonomous decision-making is serving its ultimate function on this view (i.e. as a reliable guide to what is in a patient's best interests). Our understanding of DMC to decide in a morally valuable way would then be more analogous to tightrope-walking. However, this argument would contravene the *anti-evidentiary claim.* Alternatively, one might argue that DMC should be tied to considerations of safety so that considerations of beneficence can be invoked to influence whether an individual's autonomous decision should be respected. However, this argument would contravene the *anti-hard paternalist* claim. In the absence of a revisionary argument for rejecting these claims, the analogy fails to provide a compelling case for a risk-relative standard.

I therefore believe we should reject the true-capacity threshold view. Nonetheless, I do not think we should lightly dismiss the intuitive appeal of Lawlor's 'cake or death' example. However, this is an intuition we can capture with alternative views. Prior to defending the alternative that I favour, I shall conclude this section by considering Douglas' (2022) brief defence of a risk-relative standard that pertains to the test for capacity rather than true capacity.

Recall that Douglas' draws a distinction between 'true capacity', which relates to the abilities that are necessary to make a decision that *is* autonomous, and the 'test for capacity', which pertains to the level of capacity a person must evidence for her decision to be respected *as if it were autonomous*. In a short editorial, Douglas briefly suggests a pragmatic argument for adopting an asymmetrical approach about the latter. His argument appeals to the harms involved in making different kinds of error in assessments of DMC. The first is a 'false positive' error; here, a patient is believed to have the abilities necessary to make an autonomous decision, when in fact they do not. In contrast, 'false negative' errors occur when a patient is believed to lack the abilities necessary to make an autonomous decision, when in fact they do possess these abilities. Douglas rightly notes that our decision about where to set the DMC threshold in our test for capacity will influence the proportion of each error. A more demanding standard will reduce the number of false positives, whilst a less demanding standard will reduce the number of false negatives.

The crucial premise in Douglas's pragmatic argument in favour of an asymmetrical approach to the test for capacity is that false positive assessments will be associated with greater moral costs in cases where a patient decides to refuse treatment (Douglas 2022, p. 800). The justification for this claim is that treatments offered to patients will typically be in their best interests; so, when we respect a treatment refusal after making a false positive judgment of DMC, we will fail to act in the patient's best interests; in Joe's case, we would fail to provide a safe, life-saving surgery. In contrast, when we make false positive assessments of DMC in cases where a patient is consenting to treatment, this harm does not apply; we will in fact still provide the patient with a treatment that is in their interests. As such, Douglas concludes that in cases of treatment refusal, we might plausibly be more concerned with reducing the number of false positives than we are in cases of consent. This speaks in favour of asymmetric thresholds in the test for DMC across these cases; we can thus call this the 'cost of error' view.

Douglas presents an elegant argument in favour of this interpretation of the asymmetry. However, although he rightly highlights that false positive assessments in cases of treatment refusal will involve failures to act in the patient's best interests, Douglas does not acknowledge the corresponding autonomy costs associated with false negative assessments in these cases.[2] Yet, these can be considerable; making a false negative assessment in a case of treatment refusal means that a patient's autonomous decision to *not* waive their claims against bodily interference may not be respected. Yet, even if the unwanted bodily interference is in their best interests, it can be understood to violate the underlying right if the individual was capable of autonomously waiving it, but chose not to do so. To assume that the moral costs associated with false positives are more significant than those associated with false negatives here is thus to deny the *anti-paternalist claim*; it is to deny that our moral reasons to respect autonomy should take priority over our moral reasons of beneficence. Moreover, we may note that, like false positives, the harms associated with false negatives may also be

---

[2] I have previously raised a similar objection to a view that Buchanan discusses in Buchanan (2004). See (Pugh 2020, pp. 194–195).

correspondingly higher in riskier decisions; such decisions may have a far greater bearing on the extent to which our lives proceed in accordance with our own values.[3]

As such, the theories of risk-relative standards that I have surveyed here are not compatible with key normative claims underlying DMC's gate-keeping role. In the final section, I will outline and defend a theory that is compatible with these claims.

3 The 'Quality of Evidence' View

The 'quality of evidence' view that I shall defend claims that when a decision has serious consequences, we may justifiably require higher quality evidence to inform our assessments of DMC.[4] This view shares some affinities with the 'cost of error' view. Like that view, it also suggests that the consent/refusal asymmetry pertains to the test for DMC rather than true capacity. Moreover, the justification of the view is grounded by the moral reasons to avoid errors in our assessments of DMC. However, the 'quality of evidence view' can be importantly distinguished from the 'cost of error view' as I shall now explain.

I objected to the 'cost of error' view on the basis that it is somewhat myopic; it focuses only on the moral costs associated with false positives. I suggested that we have strong moral reasons, grounded in the value of autonomy, to also avoid false negatives. Moreover, I noted that the *anti-paternalist claim* implies that avoidance of the former error should not take priority over avoidance of the latter. Here, we may also note a broader problem with seeking to avoid the moral costs associated with errors in DMC assessment by simply altering the threshold of capacity that an individual must evidence; doing so can only serve to reduce instances of one error (e.g. false positives) at the expense of increasing instances of the other (e.g. false negatives); altering the threshold is thus a somewhat blunt tool for seeking to address the moral problems associated with inaccuracy in DMC assessments (Pugh 2020, p. 196).

To avoid a myopic approach whilst also acknowledging that the costs of both kinds of error can plausibly increase in accordance with the risks of a decision, we need a more nuanced strategy that seeks to reduce the incidence of *both* errors, as the risks associated with a decision increases. The crux of the 'quality of evidence' view is that we might do so, not by raising or lowering the threshold of the capacity that a patient must evidence, but instead by requiring better quality evidence to inform our assessments of whether the relevant DMC threshold has been passed, as the risks associated with a decision increases.

To illustrate the difference between the two views, recall the example of Joe. On the cost of error view, Joe need only evidence that he passes a low threshold of capacity (T_{Low}) when he agrees to the surgery, as the welfare cost of a false positive assessment in this case is low. On the other hand, Joe would need to evidence that he passes a higher threshold of capacity (T_{High}) where he refuses the surgery, as the costs of a false positive are high.

In contrast, on the 'quality of evidence' view, the claim is that the threshold level of capacity ($T_{Simpliciter}$) that Joe must evidence passing should remain constant regardless of his decision – what should change however, is the quality of evidence that we should require to

---

[3] Consider, for example the costs of false negative DMC assessments of women refusing obstetric intervention. See Murray (2020); Zampas (2020).

[4] I first sketched the broad contours of this view in a short paragraph in Pugh (2020, p. 196). This paper is an attempt to flesh out the bare bones of that short sketch.

provide a sufficient justification of a claim about whether or not Joe surpasses ($T_{Simpliciter}$); better quality evidence aims to ensure higher *in toto* accuracy of that assessment.

The first question that this view raises is what it means to require higher quality evidence in this context. One concern here is that there is no empirical 'gold standard' of capacity assessment (Cairns et al. 2005), and the use of existing clinical tools to wholly replace clinical judgement has been subject to substantial criticism (Banner 2012). However, it would be a mistake to take this problem to undermine the more general claim that we can plausibly expect assessments of capacity to typically provide better quality evidence if they are increasingly thorough, taking into account a wider range of evidentiary sources. The thought underlying the quality of evidence view is that the provision of better quality evidence (via more thorough forms of assessment) can reasonably be expected to reduce the likelihood of *both* false negative or false positive judgments – it will, we may say, thereby increase what I shall call the *in toto* accuracy of the assessment. Indeed, if more thorough assessments did not lead to greater *in toto* accuracy, one might reasonably question the point of performing DMC assessments at all; why rely on *any* sort of DMC assessment unless we suspected that such assessments will lead to fewer incorrect verdicts of DMC than if we relied on mere guesswork or intuition alone?

To illustrate the general point further, consider the below example adapted from the Mental Capacity Act Code of Practice:

> On his daughter's recommendation and after lengthy discussion, Elliot has recently agreed to move into a care home. One day he falls and breaks his leg. In hospital, he tells staff that that he doesn't want to return to the care home. Nurses are concerned that Elliot does not understand the risks of living at home in his condition. They are also concerned that Elliot appears to be depressed and confused, and they are concerned that the deterioration in his mental health is affecting his judgment.

In the code, the purpose of the original version of this example is to illustrate when it may be appropriate for carers to seek a further professional opinion in making a DMC assessment (Mental Capacity Act Code of Practice, 2020, p.57). The code suggests that assessors could refer Elliot to a specialist in old-age psychiatry to provide an expert opinion on Elliot's mental health, and its impact on his decision-making. Depending on one's view DMC, one might add that assessors in my extended version of the case could perhaps speak to Elliot's daughter to get a better sense of his long-term views about living in a care home, and how those views relate to what Elliot is now telling them. This might enable assessors to develop an understanding of whether Elliot's current wishes are capricious or reflective of more stable values, which might be thought to betoken autonomous decision-making on some theories. Finally, they might even consider asking Elliot to undergo certain medical or psychometric tests, to provide further detail on his understanding and use of the information provided to him.

The point here is that all of these different sources of evidence could contribute to a richer and more secure analysis of the level of DMC that Elliot is able to evidence. Some evidence sources might counter grounds that could otherwise lead to a false negative; for example, the old-age psychiatrist might counter the carer's non-expert suspicion of depression. Other sources might help forestall grounds that could otherwise lead to a false positive; the daughter's evidence might suggest that Elliot's professed values at the hospital are completely inconsistent with their previous conversations, and that they should therefore be discussed with Elliot further.

Of course, there are important questions about which forms of evidence can best facilitate optimal *in toto* accuracy in DMC assessments. Crucially, the answer to these questions will depend not only on empirical matters about the objective validity of certain kinds of test, but also on normative commitments about the view of autonomy undergirding one's view of DMC (Banner 2012). For the purposes of my argument here though, the point is that requiring the acquisition of better quality evidence (via more thorough forms of investigation) in riskier decisions is a more nuanced approach to risk-relative standards because it calls for the consideration of evidence that might counter *both* false positives and false negatives. My claim is that in Elliot's case, the moral justification for performing a more thorough DMC assessment (as called for in the code of practice) is that doing so may plausibly increase its *in toto* accuracy.

This leads to a different question for the view I am presenting: Why should we reserve capacity assessments with the highest expected *in toto* accuracy only for the highest risk decisions? The answer to this is that as the thoroughness of a capacity assessment increases (to secure better quality evidence), so too will various moral costs associated with the assessment. Increasingly thorough forms of assessment will require many more resources, and there can be significant opportunity costs of this in the healthcare sector. Other costs are also plausibly borne by patients themselves. Most prosaically, some forms of assessment will require the patient to attend appointments which take up time and effort. More significantly, awareness that one's DMC is being subject to deeper scrutiny can be highly distressing. Indeed, there might plausibly be broader 'autonomy-costs' of deeper assessment. This may simply be because the additional investigation into the patient's capacity contravenes their own wishes; however, there may be deeper costs if being subject to a more thorough investigation of DMC serves to erode the self-trust of the patient, which is taken to be central to autonomy on certain views (McLeod 2002).

As detailed above, we saw that on the 'cost of error' view, the problem with reducing the number of false positives by adopting (T_{High}) over (T_{Low}) is that this strategy will unavoidably increase the number of false negatives, and their attendant moral costs. In contrast, for the 'quality of evidence' view, the problem of requiring better quality evidence to achieve greater *in toto* accuracy is that this will typically increase various other costs associated with more thorough assessments. When the harms of *both* false negative or false positive DMC assessments are high (as they plausibly are in the case in which Joe, from my initial example, refuses surgery), then the moral reasons to minimise either error can very plausibly outweigh the various moral costs associated with more thorough forms of capacity assessment. However, when the harms of *both* false negative or false positive assessments are sufficiently low (as might be the case where Joe agrees to surgery), then the moral reasons to minimise either error may plausibly be outweighed by the various moral costs associated with more thorough forms of DMC assessment. More generally, as the risk of an individual's decision increases, the moral costs of *in toto* inaccuracy will increase, and it will be easier to justify the attendant burdens of increasingly through forms of DMC assessment.

Crucially, unlike other views that I have considered in this paper, this view of the asymmetry is compatible with the three claims undergirding DMC's gate-keeping role. It is compatible with the *procedural-autonomy capacity* claim, as it accepts that DMC pertains only to abilities that are necessary for procedurally autonomous decision-making. The view merely claims that we should acquire better quality evidence of the presence of those abilities in riskier decisions. It is also compatible with the *anti-evidentiary claim*, and *the anti-hard paternalist claim;* the justification for invoking a higher standard in risky decisions neither appeals solely to considerations of beneficence, nor reduces the role of autonomy to serving as a reliable guide to what is in a patient's interests. Instead, the justification is grounded in the fact that the moral reasons associated with *both* beneficence and autonomy are likely stronger in the case of riskier decisions. Finally, the view can accommodate the intuitive pull of Lawlor's 'cake or death' example; the view suggests that whilst we might accept the DMC of the cake-eater with comparatively basic investigation, we should require the highest quality evidence (and the most thorough of assessments) to inform a judgement about the DMC of the person who instead chooses death.

I shall conclude by responding to an objection to the distinction between 'true' capacity and the 'test for capacity', a distinction that is central to both the 'cost of error' and the 'quality of evidence' views. Wilks criticised this distinction in his early defence of risk-relative standards as follows:

> If we do not assess very closely how people meet a standard it becomes much easier for them to get by without actually meeting that standard, which is in effect exactly the same as holding them to a lower standard. If I set an easier exam in a course I am not just assessing less carefully whether students meet a given standard; I am actually lowering the standard that they need to meet in order to receive a particular grade. (Wilks 1999, p. 155).

However, I believe Wilk's concern is misguided. To see why, let us reflect on the exam example. The aim of the exam is plausibly to distinguish students who have adequately learnt the class material from those who have not. Before a teacher can write an exam, she must first establish a definition of what it would be for a student to have learnt the material to an extent that warrants passing them. The aim of the test would thus be to identify students who meet this definition; call these students the 'true learners'.

Once the teacher has defined what constitutes a true learner, she must devise an exam that will distinguish the true learners, whilst avoiding misclassifications. The kinds of questions the exam involves, and the stipulated pass mark will both determine how successful the exam is in achieving this aim; I agree with Wilks in this regard. If a test has questions that are too easy, all the true learners will pass; but so too will many non-true learners. The same will be true if the passing grade is very low. However, the key point is that *neither* of these features has any bearing on the question of how we should define what constitutes a true learner. A person who passes an exam with either easy questions or a low pass grade does not thereby become a true learner, or change what constitutes true learning; instead, they become a false positive result of an exam that assesses true learning in an inaccurate (that is to say 'non-specific') manner. Accordingly, Wilks' example does not establish that there is not a meaningful difference between (i) the standards we invoke in defining a category that we want a test to identity and (ii) the standards we invoke in the test that we employ in attempting to accurately identify that category.

The importance of this distinction can be better illustrated as follows: Suppose authorities want to identify infectious people in a pandemic, so that they can quarantine those who risk harming others. To do so, they need to first establish what makes a person infectious;

suppose scientists discover that people only transmit the virus once they carry a viral load of 100 DNA copies/100 ml. Suppose, fantastically, scientists develop a test for the virus that is 100% accurate; but it can only establish whether a person has a viral load above 50 copies/ml. Should authorities use a positive result on this test as a basis for quarantining an individual? The answer will depend on (i) the strength of the moral reasons to identify the truly infectious (ii) the likely number of people who will test positive but be *incorrectly* classified as truly infectious (i.e. those with a viral load between 50 and 99 DNA copies/ml), and (iii) the costs of such misclassifications (such as unnecessarily restricting liberty). Balancing these considerations requires a complex value judgment; but that is a quite separate issue to the prior *scientific* question of how much viral load a person must *actually* carry to be truly infectious.

It is true that using the test to identify who should be quarantined would, in effect, amount to treating anyone with a viral load above 50 copies/ml *as if* they were truly infectious; why not follow Wilks in claiming that this is just to say that such individuals are truly infectious in practice? This suggestion overlooks something important; discrepancies between the two categories can alter how and whether our actions are morally justified. To illustrate, suppose we initially believed that the benefits of using the test would be outweighed by the moral costs of misclassifications. However, imagine now that the virus mutates; scientists establish that people now become truly infectious once they carry a viral load of just 50 copies/ml. In these new circumstances, the test would not lead to *any* misclassifications.

This makes a considerable moral difference; the moral costs associated with misclassifications no longer obtain, so they cannot outweigh the benefits of using the test. However, it is hard to see how this difference could matter on Wilks' view. Both before and after the mutation, we are considering the permissibility of treating those with a viral load of 50 copies/ml as if they were truly infectious; for Wilks, in both cases, permitting such treatment amounts to saying that these individuals are truly infectious in practice. Now, unless we appeal to the importance of the distinction between standards of being truly infectious and standards we set in testing for infectiousness, it is hard to see how Wilks could account for the plausibility of the claim that the mutation affects whether the use of the test is justifiable. Similarly, without distinguishing true capacity and the test for capacity, it is not clear that we can account for the importance of the difference between (i) the moral reasons to respect a decision we know to be truly autonomous and (ii) the moral reasons to respect a choice *as if it were* autonomous, under conditions of uncertainty.

4 Conclusion

The 'quality of evidence' view I have outlined is an improvement on existing accounts of risk-relative standards for three reasons: (i) it is compatible with key normative claims underlying the gate-keeping role of DMC, (ii) it accommodates the intuitive pull of Lawlor's cake or death case and (iii) it is sensitive to the fact that the moral costs of different errors in assessing DMC increase in accordance with risk. Accordingly, I believe that it offers the strongest defence of a risk-relative approach to DMC, and that such standards are not nonsensical.

However, it is important to be clear about the limited implications of this argument. As Lawlor astutely highlights, we can identify a number of superficially alike but actually quite different asymmetries that are invoked in practical debates about risk-relative standards of DMC. These include asymmetries with respect to (i) the choices that are offered to patients in clinical encounters, (ii) whether tokens of consent and refusal are honoured, and (iii) legal rights (Lawlor 2016). My view does not straightforwardly entail any of these further asymmetries; all it entails is that we have strong moral reasons to obtain better evidence of DMC in the case of riskier decisions, on the assumption of DMC's gate-keeping role. The question of how we ought to *act* on such an assessment of DMC can only be answered once we have taken a much wider range of considerations into account.

That said, the quality of evidence view has an important theoretical implication for risk-relative standards. On this view, the plausibility of such standards ultimately depends on a nuanced understanding of what constitutes the sort of high quality evidence that can be expected to increase the *in toto* accuracy of DMC assessments. Insofar as this understanding must first be grounded in a robust account of which abilities are required for autonomous decision-making, philosophical discussions of risk-relative standards of DMC could usefully reorient their focus towards questions about epistemology and the nature of autonomy, as well as questions about evaluative conflicts between autonomy and beneficence.

Acknowledgements I would like to thank Joanna Demaree Cotton and the audience at the European Society of Philosophy of Medicine and Healthcare Conference 2023 for their invaluable feedback on earlier drafts of this paper.

Author Contributions Sole authored.

Funding This research was funded by The British Academy (KF8\230096).

Data Availability N/A

Declarations

Ethical Approval N/A

Informed Consent N/A

Statement Regarding Research Involving Human Participants and/or Animals N/A

Competing Interests N/A

References

Banner NF (2012) Unreasonable reasons: normative judgements in the assessment of mental capacity. J Eval Clin Pract 18(5):1038–1044. https://doi.org/10.1111/j.1365-2753.2012.01914.x

Beauchamp TL, Childress JF (2001) Principles of biomedical ethics, 5th edn. Oxford University Press

Birchley G (2021) The theorisation of 'best interests' in bioethical accounts of decision-making. BMC Med Ethics 22(1):68. https://doi.org/10.1186/s12910-021-00636-0

Bock GL (2012) *Jehovah's Witnesses and autonomy: Honouring the refusal of blood transfusions*. 652–656

Buchanan A (2004) *Mental capacity, legal competence and consent to treatment*. 415–420

Buchanan AE, Brock DW (1989) Deciding for others: the ethics of surrogate decision making. Cambridge University Press

Buller T (2001) Competency and risk-relativity. Bioethics 15(2):93–109. https://doi.org/10.1111/1467-8519.00218

Cairns R, Maddock C, Buchanan A, David AS, Hayward P, Richardson G, Szmukler G, Hotopf M (2005) Reliability of mental capacity assessments in psychiatric in-patients. Br J Psychiatry 187(4):372–378. https://doi.org/10.1192/bjp.187.4.372

Cale GS (1999) Risk-related standards of competence: continuing the debate over risk-related standards of competence. Bioethics 13(2):131–159. https://doi.org/10.1111/1467-8519.00137

Cholbi M (2017) Paternalism and our rational Powers. Mind 126(501):123–153. https://doi.org/10.1093/mind/fzv205

Culver C, Gert B (1990) *The Inadequacy of Incompetence*. 619

Douglas T (2022) Pragmatic argument for an acceptance-refusal asymmetry in competence requirements. J Med Ethics 48(11):799–800. https://doi.org/10.1136/jme-2022-108662

Dworkin R (1993) Life's dominion: an argument about abortion and euthanasia. HarperCollins

Faden RR, Beauchamp TL (with, King N (1986) M. P.). *A history and theory of informed consent*. Oxford University Press

Feinberg J (1984) The moral limits of the criminal law. Oxford University Press

Fogal D, Schwan B (forthcoming). Ditching decision-making capacity. J Med Ethics. https://doi.org/10.1136/jme-2023-109047

Friedman M (2003) Autonomy, gender, politics. Oxford University Press

Harris J (2003) Consent and end of life decisions. J Med Ethics 29(1):10–15. https://doi.org/10.1136/jme.29.1.10

Lawlor R (2016) Cake or death? Ending confusions about asymmetries between consent and refusal. J Med Ethics 42(11):748–754. https://doi.org/10.1136/medethics-2016-103647

Manson NC (2015) Transitional paternalism: how Shared Normative Powers give rise to the asymmetry of adolescent consent and refusal. Transitional Paternalism Bioeth 29(2):66–73. https://doi.org/10.1111/bioe.12086

McLeod C (2002) Self-trust and reproductive autonomy. MIT Press

McMahan J (2002) The ethics of killing: problems at the margins of life. Oxford University Press

Mental Capacity Act Code of Practice. (2020), October 14 GOV.UK. https://www.gov.uk/government/publications/mental-capacity-act-code-of-practice

Mill JS (2003) *On Liberty* (D. Bromwich & G. Kateb, Eds.). Yale University Press. http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=3420105

Murray C (2020) Troubling consent: Pain and pressure in Labour and Childbirth. In: Pickles C, Herring J (eds) Women's birthing bodies and the law: unauthorised intimate examinations, power, and vulnerability. Hart Publishing, an imprint of Bloomsbury Publishing

Pickering NJ, Newton-Howes G, Walker S (2022) Risk-related standards of competence are a nonsense. J Med Ethics. https://doi.org/10.1136/medethics-2021-108107

Pugh J (2020) Autonomy, rationality, and contemporary Bioethics. Oxford University Press

Savulescu J, Momeyer RW (1997) *Should informed consent be based on rational beliefs?*

United Nations (2006) *Convention on the Rights of Persons with Disabilities*. https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/convention-on-the-rights-of-persons-with-disabilities-2.html

Wicclair M (1991) *Patient decision-making capacity and risk*. 91–104

Wilks I (1999) Asymmetrical competence. Bioethics 13(2):154–159. https://doi.org/10.1111/1467-8519.00139

Zampas C (2020) Human rights and gender stereotypes in Childbirth. In: Pickles C, Herring J (eds) Women's birthing bodies and the law: unauthorised intimate examinations, power, and vulnerability. Hart Publishing, an imprint of Bloomsbury Publishing