



UNIVERSITY OF LEEDS

This is a repository copy of *Advance Directives: The Principle of Determining Authenticity*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/230728/>

Version: Accepted Version

---

**Article:**

Carter, M. [orcid.org/0000-0003-0683-3874](https://orcid.org/0000-0003-0683-3874) (Cover date: January/February 2022) *Advance Directives: The Principle of Determining Authenticity*. The Hastings Center Report, 52 (1). pp. 32-41. ISSN: 0093-0334

<https://doi.org/10.1002/hast.1338>

---

This is the peer reviewed version of the following article: Matilda Carter, "Advance Directives: The Principle of Determining Authenticity," *Hastings Center Report* 52, no. 1 (2022): 32-41. DOI: 10.1002/hast.1338, which has been published in final form at <https://doi.org/10.1002/hast.1338>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## Advance Directives: The Principle of Determining Authenticity

### **Abstract**

There is a well-established debate about the moral validity of advance directives over people living with dementia in bioethics, which is often cast as a clash between two principles: respecting autonomy and beneficence towards patients. In this paper, I adapt John Christman's account of authenticity as non-alienation to bypass this clash of principles, instead arguing that there need be only one principle in substitute decision-making: the principle of determining authenticity.

### **Introduction**

While studying at medical school, neurosurgeon Andrew Firlik met a woman named Margo who lived with Alzheimer's disease. Margo painted the same picture every day, read through her mystery novel in a seemingly random order and did not appear to be able to remember people's names. Contrary to his initial intuitions about this kind of life, Firlik described her as "undeniably one of the happiest people [he has] ever known."<sup>1</sup> In an influential chapter of his book *Life's Dominion*, Ronald Dworkin asks us to imagine that Margo had signed an advance directive that, in the event she developed dementia, she not be provided with life-saving treatment.<sup>2</sup> Were she to come to need it, we would be confronted with a dilemma: do we allow Margo to die even though she is plainly enjoying her life?

Two candidate principles are invoked in medico-legal contexts to determine the best course of action when a patient is deemed unable to make a decision for themselves: the *principle of respect for autonomy* and the *principle of beneficence towards patients*.<sup>3</sup> The former favours a *substituted judgment standard*, wherein the subjective viewpoint of the patient is reconstructed, with the aid of an advance directive where possible, to enact a decision they would have made for themselves. For patients who are temporarily incapacitated, this is

thought to be the appropriate principle.<sup>4</sup> The latter principle favours a *best interests standard* which takes a more general view of interests, and is thought to be more appropriate for patients who have never been rationally autonomous.<sup>5</sup>

Which principle to invoke in cases like Margo's, however, is the subject of dispute, because she used to possess rational competence, but her capacities have permanently declined. Some, like Dan Brock, argue that we should follow the principle of respecting autonomy and uphold advance directives in all such cases, as they represent autonomous choices made before such choices were no longer possible.<sup>6</sup> Another plausible view holds that, because Margo is no longer the sort of being to whom respect for autonomy is owed, the principle of beneficence applies. Dworkin, however, bypasses the debate by arguing that both the principle of respecting autonomy and the principle of beneficence recommend upholding Margo's advance directive.<sup>7</sup>

In this paper, I argue that there need only be one principle in substitute decision-making: the principle of determining authenticity. This principle favours a substituted judgment standard in all cases and instructs decisionmakers to determine what the patient would authentically prefer to happen—not merely in light of their past decisions, but in light of their present settled dispositions. Adhering to this principle entails that, in a significant range of cases, an advance directive can (and, indeed, ought to) be overruled.

In section 1, I outline Dworkin's arguments from autonomy and beneficence and demonstrate that they both rely on the same, integrity-oriented account of authenticity. In exploring how it can be defended against later criticisms, I argue that an embryonic version of the view I wish to defend can be derived from Dworkin's work. However, in section 2, I dispute the account of authenticity underlying Dworkin's arguments and, instead, adapt John Christman's alternative view of authenticity as non-alienation to argue that Margo continues to hold authentic preferences which can be determined by decisionmakers. Having established the principle of determining authenticity I argue that, in simple cases, in which the person's

life is weighed against a painless life-saving intervention, the advance directives of Margo and other people living with dementia ought to be overruled. In section 3, I explicitly ground the principle of determining authenticity in the theory of freedom as non-domination in order to lay out how to come to decisions in complex cases and defend the use of (defeasible) advance directives.

The arguments of this paper cohere with a growing acceptance of the moral weight of present-day preferences in the decisionally-impaired<sup>8</sup> and of the importance of authenticity in surrogate decision-making.<sup>9</sup> The key contributions of this paper, however, are twofold. First, I specify a particular account of authenticity, and use it to argue that respect for authentic preferences ought to replace the autonomy/beneficence paradigm. Secondly, I offer a methodological pathway for determining authentic preferences and coming to a decision which respects them.

## **1. Dworkin and the Value of Authenticity**

Dworkin's contribution to the debate on substitute decision-making for people living with dementia is significant, because he dispels the central tension in the case: the conflict between the past and present self. On his account, whether we think in terms of respecting her autonomous choices or, instead, acting in her best interests, we will arrive at the same conclusion; either way, Margo's advance directive ought always to be upheld. In this section, I argue that Dworkin reaches the same answer because he is tacitly committed to the view that both principles ask the same question: what decision best reflects her authentic self? An embryonic version of the view I propose is present, then, but later I will argue that it reaches a mistaken conclusion because it is based on an indefensible view of authenticity.

### *1.1 Dworkin's Argument from Autonomy*

Following other scholars who have discussed the subject, Dworkin argues that respecting autonomy requires executing Margo's advance directive, even though it seems apparent from Firlik's account that she is consistently happy.<sup>10</sup> Accepting that it strikes many as counterintuitive to overrule a person's present wishes in favour of those expressed in the past, Dworkin offers an analysis of the grounds for respecting autonomous decisions to persuade the reader of his view.<sup>11</sup>

Dworkin labels one plausible account of these grounds the *evidentiary view*, which holds that "we should respect the decisions people make for themselves, even when we regard these decisions as imprudent, because each person generally knows what is in his own best interests better than anyone else."<sup>12</sup> If we accept this view, he argues, we should not extend the principle of respecting autonomy to people living with dementia, because they often act in ways which indicate that they do not know what their best interests are. However, Dworkin concedes, given the existence of persistent smokers, it seems likely that this is also true of many people living without dementia.<sup>13</sup> Thus, he rejects the evidentiary view, since it relies on a demonstrably false empirical claim.

If we follow Dworkin in rejecting the evidentiary view, the value of autonomy cannot be derived from its ability to protect our wellbeing. Instead, he proposes that respecting autonomy involves the protection of "people's general capacity to lead their lives out of a distinctive sense of their own character, a sense of what is important to and for them."<sup>14</sup> So understood, we do not intervene to prevent a person from smoking, despite the fact that we may think we know it is against their best interests. Indeed, on this view, respecting autonomy *demand*s allowing people to take decisions that we believe are against their best interests, because we respect their capacity to live a life which is distinctive to them.

As, according to Dworkin, this is not a capacity that people living with dementia in the advanced stages possess, however, respecting Margo's autonomy demands respecting her *most*

*recent* expression of this capacity: the advance directive. He concludes that “it is no kindness to allow a person to take decisions against his own interests in order to protect a capacity he does not and cannot have.”<sup>15</sup>

Dworkin calls his preferred account the “integrity view of autonomy” and it is fully realised when a life displays “overall integrity and authenticity.”<sup>16</sup> Although many may not achieve this, he argues that respect for autonomy is intended to protect the capacity we all have to reach it, by allowing us to choose “how far and in what form” we will seek this aim.<sup>17</sup> It is unclear, from the argument presented, if Dworkin sees integrity and authenticity as synonymous, but it seems clear the goal of the integrity view of autonomy is to respect our ability to be true to ourselves in some important respect. We respect Margo’s autonomy in executing the advance directive, on this view, because when she was last capable of thinking about the kind of life she wanted to live, she chose not to receive life-saving treatment.

### *1.2 Dworkin’s Argument from Beneficence*

The advance directive, on the integrity view of autonomy, is the final expression of Margo’s capacity to pursue a life that is true to herself. However, Dworkin acknowledges that the idea that we should not try to save the life of someone who is plainly content is intuitively troubling.<sup>18</sup> While remaining firm that refusing to execute the advance directive disrespects Margo’s autonomy, he acknowledges that some may see a conflict between what the principle of beneficence requires and what respect for autonomy requires.<sup>19</sup> Dworkin, however, dismisses this conflict by arguing that it is not in Margo’s best interests to continue to live.<sup>20</sup>

Dworkin’s analysis of the principle of beneficence divides interests into two categories. On the one hand, people have *experiential interests*; things we do “because we like the experience of doing them.”<sup>21</sup> On the other, they have *critical interests*: “interests that it does make their [lives] genuinely better to satisfy.”<sup>22</sup> While both are important, it is the latter that

represent critical judgments about the shape their lives should take and the kind of choices that are “not only good at the moment but in character for them.”<sup>23</sup> Acting in someone’s best interests, according to Dworkin, is acting in light of their critical interests, as only they reflect the capacity to pursue a life of integrity and authenticity.<sup>24</sup>

Margo, evidently, has experiential interests, but Dworkin argues that she has “lost the capacity to think about how to make [her] life more successful as a whole.”<sup>25</sup> As such, Dworkin argues, she has “no contemporary opinion about [her] critical interests.”<sup>26</sup> She still possesses them, however, because critical interests are necessarily evaluative about a whole life; the advance directive is an expression of Margo’s opinion on her critical interests while she still had the capacity to conceive of them. There is no conflict between respecting Margo’s autonomy and the principle of beneficence, then, because when considering a person living with dementia’s best interests we should “consider how the fate of a demented person can affect the character of his life.”<sup>27</sup> This entails acting in a way that is compatible with her critical interests, even if it conflicts with her experiential interests, as, according to Dworkin, it is only in disrespecting the former that we fall short of what respect for persons requires.<sup>28</sup>

The reason for this apparent lack of conflict is that both of Dworkin’s arguments appeal to the same idea: that the capacity to pursue a life of overall integrity and authenticity must be protected. In both cases, Dworkin asks us to consider which decision would be concordant with Margo’s character, preferences and values. As, on his view, she is no longer capable of making such a judgment, we must respect the judgment she made in the advance directive. Although he does not say so explicitly, we can infer that the reason why both principles lead to the same answer is that both principles ought to ask the same question: *what decision best reflects Margo’s authentic self?* As Margo, on this view, no longer possesses the capacities necessary to make a judgment on what her authentic self consists of, we should respect the judgment made in the advance directive.

### *1.3 Jaworska's Response from the Capacity to Value*

Both of these arguments face an important objection which Dworkin can only respond to in a way that illuminates his contestable view on authenticity. Drawing on empirical accounts of people living with Alzheimer's disease, Jaworska argues that even those with advanced dementia maintain a capacity to value, in the sense of being capable of holding something to be important to them.<sup>29</sup> She argues that this insight reveals flaws in both of Dworkin's arguments and, as such, we may have reasons (in particular circumstances) to overturn advance directives.

Regarding beneficence, she argues that Margo is not incapable of conceiving of what is important to her. Although she concedes to Dworkin that these values do not arise from critical reflection over a person's whole life narrative, she argues that these values represent a *time-specific* sense of self.<sup>30</sup> Jaworska argues that this capacity to reflect on what is important to us in the moment can be distinguished from mere desiring through three features: we think we are correct in our values; those values are tied up with our sense of self-worth; and the importance of realising them is independent of our experiences.<sup>31</sup> As such, the capacity to value should be all that is required to generate a contemporaneous opinion on our critical interests.<sup>32</sup>

Regarding respect for autonomy, Jaworska argues that people living with dementia can still make autonomous decisions. This is because Dworkin ties autonomy to the capacity to express our character, and people in the advanced stages of dementia remain capable of generating a present sense of this character; thus we have reason to respect their decisions as autonomous.<sup>33</sup> While she concedes that Margo and others like her may need assistance to put the values, preferences and commitments they possess into practice, she argues that this does not mean they have forfeited the right to respect for their autonomy.<sup>34</sup> Having argued that dementia does not automatically discount the capacity to value, Jaworska concludes:



Thus, a possibility opens up that the capacity for autonomy ought not to be thought of as the capacity to carry out one's convictions into action without external help, a capacity that requires reasoning through complex sets of circumstances to reach the most appropriate autonomous decisions; rather, that the capacity for autonomy is first and foremost the capacity to espouse values and convictions, whose translation into action may not always be fully within the agent's mastery.<sup>35</sup>

If these two compelling arguments hold, respecting Margo's autonomy and treating her with beneficence would recommend overruling the advance directive. However, although Dworkin must concede that people living with dementia are capable of valuing, he need not concede that these apparent values are constitutive of a person's critical interests nor worthy of respect on the integrity view of autonomy. After all, Dworkin's arguments about what the principles of autonomy and beneficence require are not grounded on the idea that we should respect *any* expression of value. Rather, he is explicit that respect for persons requires only that we respect those values that arise from the expression of a particular capacity for critical self-evaluation.

In one formulation, he describes the relevant capacity as that which allows us to act "out of a *genuine* preference or character or conviction or sense of self" (emphasis mine).<sup>36</sup> Dworkin, then, implicitly allows for the possibility that people living with dementia may appear to have a sense of self, or to possess preferences, convictions or values. Indeed, embedded in this formulation is the idea that sometimes these mental elements are not genuine. Thus, a plausible Dworkinian response to Jaworska is to argue that expressions of value from persons incapable of exercising this capacity are not genuine and, thus, are not constitutive of what respect for persons requires.

The looseness of Dworkin's language here is unhelpful so, for exegetical purposes, it can be inferred that Dworkin here is using the word *genuine* in a way that is synonymous with *authentic*. Likewise, his formulation of the capacity to pursue a life of overall integrity and authenticity can be reduced to the capacity to pursue overall integrity, as he does not distinguish

between the two concepts. Doing so clarifies a plausible Dworkinian response to Jaworska's criticism: that respect for persons only requires that we respect authentic (genuine) values, and the only authentic values are those that are oriented towards pursuit of a life of overall integrity (narrative coherence of character, commitments and long-standing values).

Although this Dworkinian defence might appear to be employing a kind of no true Scotsman-type fallacy, it is more intuitive than it first appears. After all, if I were to express a set of values while under the influence of a powerful hallucinogenic drug which conflicted sharply with my sense of character before my state of mind was altered, it would be quite strange to think of these as authentic. Certainly, I would not want to be held to these expressions of value if I later came to be free of the mind-altering influence. On Dworkin's *integrity-oriented* account, the reason why these values would be inauthentic is that, because of the influence of the drug, my capacity for living a life that is authentically mine through critical self-reflection would have been hampered. Dementia too has mind-altering effects, so the idea that it might lead to inauthentic expressions of values is not so far-fetched.

If Dworkin were to take this route, then he would be moving further away from the autonomy-beneficence paradigm. The case for not respecting Margo's present wishes is no longer that they do not represent *any sense of character*, but that they do not represent her *authentic character*. Respecting autonomy or acting in her best interests involve exactly the same thing; respecting a preference which is authentic. It seems possible, then, to derive a new principle for substitute decision-making from Dworkin: the principle of determining authenticity.

Whether or not Margo's present wishes should be respected, then, turns on whether or not they are authentic mental elements. The account of authenticity that leads Dworkin to reject present-day expressions of values, however, is disputable. Arguing that people living with

dementia should be able to overrule advance directives, as I intend to, requires disputing Dworkin's *integrity-oriented account of authenticity* and offering a coherent alternative.

## **2. Determining Margo's Authentic Mental Elements**

In the previous section I argued that, in order to meet the challenge from Jaworska, Dworkin must commit himself to the view that we ought only to respect authentic mental elements as defined as those developed through the use of the capacity to live a life of overall integrity. In this section, I reject this view of authenticity by showing that reaching an overall integrity of character is neither possible nor desirable. As an alternative, I adopt a modified form of John Christman's view of authenticity as non-alienation to demonstrate that, in simple cases, we ought to infer that people like Margo have an authentic preference to receive life-saving treatment, which ought to be respected.

### *2.1 Problematising the Integrity-Oriented Account of Authenticity*

The integrity-oriented account of authenticity is predicated on the view that all cognitively-able people possess a capacity to reach a life of overall integrity in terms of a narrative coherence of character. Although Dworkin is explicit that many of us will never reach it, it needs to be an achievable goal to be a genuine capacity that we all possess. We ought to be suspicious of this line of reasoning, however, because narratives we set for our lives can be disrupted in multiple, unwanted ways. Severe illnesses, family tragedies and global events can change the character of our lives dramatically and, because the future is unpredictable, we cannot usually plan the shape of our lives in advance. Indeed, as R. L. Berghmans argues, this uncertainty is a key problem with drawing up strict advance directives, as we are also unable to fully anticipate how events may change the character of our lives before they happen.<sup>37</sup>

Moreover, sometimes we find that things we once considered valuable or integral to our characters lose their appeal to us as our lives change and we engage with new circumstances and new information. When I was 18, I was the singer and main songwriter for a band, and I was certain that my life would be incomplete if I never had a chance to pursue music as a career. Many years later, with reflection and with a newfound ability to notice how bad our music was, I have jettisoned this belief and no longer think pursuing such a career is integral to my character. Would 18-year-old-me be appalled at this betrayal of my integrity? Probably, but this indicates that emphasising a capacity to reach an overarching integrity of character ignores the fact that, for most of us, our character changes significantly as we age and engage with new experiences and information. Indeed, Rebecca Dresser argues against Dworkin's distinction between critical and experiential interests for very similar reasons; many of us take life one day at a time, and the division between our critical projects and our experiential enjoyment is fuzzy.<sup>38</sup>

A Dworkinian, of course, could respond that changes in character over time are part of what it means to reach an overall integrity of character. I can make sense of my desire to grace the front cover of rock magazines retrospectively, by tracing the evolution of the values and preferences I hold over time. The problem with this much looser notion of narrative consistency, however, is that it is difficult to see how anyone could fail to meet this goal; everything that happens to us makes us who we are. Dementia poses no threat to this kind of narrative consistency, because it is just a further evolution of the story; Margo's happiness while living with dementia is just as much of a change in her character as my abandonment of my dreams of success in music, not a betrayal of its integrity.

Indeed, this is the kernel of the dispute between Jaworska and Dworkin: whether or not the values that people living with dementia generate reflect their authentic characters. If narrative coherence refers only to being able to trace the evolution of character, values and

preferences over time, then there is nothing narratively incoherent about a person's character changing due to the development of dementia. To deny the authenticity of the values Jaworska argues people living with dementia are capable of possessing, then, Dworkin must appeal to the stronger account of narrative coherence of character, with all of its implausibility attached. Even were we to grant that it is possible to achieve such a goal however, it does not follow that pursuing it and, by extension, protecting the pursuit of it, are valuable.

From a psychological perspective, there is something corrosively narcissistic about the pursuit of overall narrative consistency. There is a growing recognition in therapeutic practice that many mental health issues are exacerbated by an overidentification with life narratives and an essentialising of character.<sup>39</sup> The process of repeatedly telling themselves the same story about their lives and their character has been found, for instance, to exacerbate suicidal ideation in people with symptoms of borderline personality disorder.<sup>40</sup>

Moreover, it is quite plausible to think that there are circumstances in which our preferences, values and commitments *should* change. We would be unlikely to think, for instance, that white supremacists are losing something of value if they jettison their ideological commitments when confronted with data disputing pseudo-scientific accounts of racial hierarchies. If the dedicated pursuit of an overarching integrity of character carries a risk of rigidity and pathology, then, Dworkin's exaltation of it is, at the very least, questionable.

A Dworkinian might respond by arguing that the integrity view does not exalt a life that is rigidly consistent in character, but this misses the point: the psychological evidence suggests that using overarching narratives to evaluate our lives and making categorical statements about who we are inhibits our capacity for growth. Moreover, it may also inhibit our ability to recover from trauma; therapeutic professionals often use a form of narrative therapy to help survivors of sexual abuse and violence to construct a new sense of themselves.<sup>41</sup> Adherence to an

integrity-oriented account of authenticity requires us to view such techniques as a process of creating inauthentic mental elements, which seems wildly implausible.

As the pursuit of an overarching integrity of character is implausible, probably impossible and in any case undesirable, Dworkin's integrity-oriented account of authenticity ought to be jettisoned. That they are disconnected from a broader sense of narrative coherence, then, does not seem enough to dismiss Margo's present wishes. Nevertheless, there is still something intuitive about the idea that authentic preferences, desires and wishes should carry more weight than those that are inauthentic; Dworkin's claim that we should refuse a blood transfusion to a Jehovah's witness begging for one in a delirious state, when we are certain he will deeply regret receiving it upon recovery seems right.<sup>42</sup> It is worth considering, then, whether an alternative account can undergird and bolster the principle of determining authenticity.

## *2.2 Authenticity as Non-Alienation*

John Christman, as part of a broader argument about autonomy, argues that authentic mental elements are those that a person would not deeply repudiate upon a hypothetical or actual reflection over a variety of circumstances.<sup>43</sup> This process is engaged in in order to determine, "the characteristic mode of thinking and moving that the self-schema embodies," and to derive a person's "settled character."<sup>44</sup> *Authenticity as non-alienation*, then, jettisons the notion of connection to a wider life narrative in favour of the settled disposition of a person towards their own mental elements.

Of course, Margo does not have the capacity to reflect over a variety of circumstances, so it might be thought that her current preferences cannot be considered authentic. Indeed, it is unlikely that Christman intended his account to be used to defend the view that we ought to treat present expressions of mental elements by people living with dementia as authentic; while

expounding his argument he explicitly states that to be authentic on this account requires the capacity to “reflect adequately without constriction, pathology, or manipulation.”<sup>45</sup> However, he also says that the reflection need not actually take place,<sup>46</sup> which raises the possibility that this capacity need not actually be present.

Christman rightly points out that many of our mental elements are unchosen and requires only that we would not be alienated from them *if* we were to reflect on them.<sup>47</sup> Given the reflection is hypothetical, it does not seem necessary that the person in question actually requires underlying rational capacities. Thus, we can use evidence of Margo’s disposition, stated preferences and behaviour to engage in this reflection on her behalf and come to a conclusion about what she would choose if she were capable. Indeed, empirical evidence suggests that it is common for decisionally-impaired patients to retain a set of stable preferences.<sup>48</sup> This is certainly true of Margo, who engages in the same tasks and expresses the same contentment every day. Thus, the task for the decision-maker is to determine whether or not she would be deeply alienated from these stable preferences if she were capable of reflecting on them.

There are two ways of understanding this ‘if’. One is that the relevant counterfactual is Margo before this capacity to reflect and decide had declined. In this case Margo would clearly repudiate her present preferences, because we would be winding back the clock to the time that Margo had signed the advance directive. However, this Margo is missing information which is central to the decision being made. As raised by Dresser, before onset, Margo would not have been aware of the experience of living with dementia nor of the therapeutic options that would be available to her once it developed.<sup>49</sup> Moreover, as Emily Walsh argues, dementia is a cognitive transformation, which changes the way a person views themselves.<sup>50</sup> As such, how Margo would have before developing dementia is not a reliable indicator of what she would conclude now, were she to have the capacity to reflect.

The relevant counterfactual for present-day Margo, then, is present-day Margo with the added capacity to reflect on the advance directive and the preferences she now holds. When deciding whether or not to execute the advance directive, determining whether or not Margo would be alienated requires engaging in this reflection for her. Given what we know of her, that she is the happiest person Firlik has ever met and that this contentment was apparent to him on multiple occasions, it seems as if Margo would be highly unlikely to repudiate her clear contentment with her living situation and would be much more likely to be alienated from the preferences stated in the advance directive.

A sceptic might respond, however, that the variety of imagined points of reflection must include circumstances in which she does not have dementia. Such a claim, however, would profoundly misunderstand dementia and the point of this reflective process. This is a condition that is typified by progressive cognitive decline for which there is no cure; in every possible circumstance where we imagine Margo reflecting on her desire to live, she would have dementia. There is little point in considering how Margo would feel in impossible counterfactuals, because there is nothing at stake: this deep sense of alienation that accompanies an inauthentic desire would never come to pass.

A more nuanced critique might be to argue that the very addition of the capacity to reflect might change Margo's preference set. So understood, we are taking Margo out of one of the effects of her dementia and asking her to decide whether she wants to carry on living with it once the reflective process is over. Here Margo might become influenced by the kind of narrative concerns that underline Dworkin's argument from narrative coherence, and although we have no evidence that she would be deeply alienated from her experiential pleasure, she might suddenly decide she is concerned about how living with dementia will affect the character of her life.



That this is conceivable must be granted, that it is relevant, from the evidence we have, however, need not. Inevitably, there is some epistemic uncertainty involved in engaging on a reflective process from the outside and decisionmakers will need to avoid becoming too drawn into a discussion about how added capacities might affect preference sets. As Margo will never cease to have dementia, any conceivable concern about narrative coherence that she does not currently possess but might arise from the magical granting of reflective prowess is moot. The decisionmaker is only required to externally reflect on the evidence they have in front of them. Reflecting on a variety of circumstances requires considering how the removal of temporary internal or external factors might affect apparent preferences, based on the evidence that is available, in order to determine a settled disposition.

What is not required, however, is to consider hypothetical preference-sets for which there is no evidence Margo holds in the present-day. Perhaps if there were some indication of present-day existential angst or other concerns that cohered with those presented in the advance directive, then there would be room for a reasonable dispute over which interests are alienating. In the case as presented by Firlík, however, Margo's contentment and preference to continue living is stable and unchallenged.

This feature of the view is important in differentiating it from cases of addiction or cases of temporary states from which a person would recover. For instance, if an alcoholic gives the key to his wine cooler to his partner and asks her not to give it back to him, even if he asks, a craving may not be an authentic mental element because it is one that we have good reason to believe he will repudiate with that craving removed. Of course, it might be the case that the alcoholic would deeply repudiate his intention to quit drinking while he is experiencing a craving but, if this is the case – if he is truly alienated – then we have no choice but to conclude that his intention to quit is not yet a settled part of his character.

Likewise, in Dworkin's Jehovah's Witness example, a single instance of begging for a blood transfusion in a delirious state, is not an expression of an authentic preference because if the person reflected on this claim, with the temporary delirium removed, we would have good reason to believe he would be alienated from it. If, however, the Jehovah's Witness has shown evidence of questioning or feeling alienated from his faith, then we may have reason to think that the faith is not part of his settled disposition.

Although my tentative conclusions to these two cases may be disputable, it ought to be clear that most forms of dementia are unlike this: Margo's dementia is permanent, so the series of circumstances we are imagining her reflecting in must all feature the dementia. Thus, in a simple case in which the life-saving treatment poses no risk to her physical health and her advance directive stipulates only that she does not want to live with dementia, Margo's persistent happiness is evidence of a settled disposition to carry on living which should be respected. The advance directive, then, does not respect her settled disposition, meaning it represents inauthentic preferences and can be overruled.

In sum, Dworkin's account of authenticity as orientation toward integrity is untenable and cannot be used to undergird the principle of determining authenticity. I have argued, instead, that an adapted version of Christman's account of authenticity as non-alienation is preferable and capable of helping us to determine what a person's authentic preference is. I have concluded, then, in simple cases in which there is no risk of harm to Margo in the course of the treatment, the advance directive ought to be overruled. Nevertheless, it would be unwise to dispose of advance directives entirely. Indeed, there is a significant range of cases in which they are vital for this external reflective process.

### **3. Maintaining (Defeasible) Advance Directives**

Following an adapted version of Christman's account of authenticity as non-alienation, the principle of determining authenticity, in simple circumstances, recommends overruling Margo's advance directive. A logical extension of this would be to argue against the use of advance directives in any circumstances. In this section, however, I defend their use, so long as they are understood as defeasible, in order to meet the complexities of real-life medical practice.

In order to appraise the value of advance directives in complex scenarios, it is necessary to gain conceptual clarity on what they are and why they are useful. A crucially under-analysed element of the advance directives debate is the relationship between the decisionmaker and the person in need of treatment. When considering the morality of providing medical treatment when a person is deemed not to have the capacity to make a decision for themselves, we must never lose sight of the fact that we are talking about someone making a decision for someone else.

The principle of determining authenticity, as I have described it here, is grounded in a specific conception of freedom. Rather than freedom from interference, which is implied by the principle of respecting autonomy, respect for authenticity is grounded in the idea of freedom as non-domination. So understood, a person is free as long as they are not exposed to the arbitrary will of another.<sup>51</sup> Key to understanding this conception of freedom, is understanding that it does not rule out interference, as long as that interference tracks a person's own interpretation of their interests.<sup>52</sup>

As decisionally-impaired patients require someone else to make medical decisions for them, they are vulnerable to this kind of threat to freedom. Without an understanding of what that person's own interpretation of their interests is, medical decisions may be made that do not track them, which represents a violation of their liberty. Advance directives, insofar as they

stand as statements of these interests, can act as a guard against this, by making it clear to decisionmakers what the person's own, authentic interpretation of their own interests is.

Nevertheless, the simple version of the case of Margo demonstrates that executing an advance directive may not always respect what the person would authentically wish to happen, particularly in cases of dementia. Despite the threat of domination, the view that advance directives are the sole authority on a person's authentic wishes or preferences remains mistaken. Rather, it is better to think of them as expressions of preferences which can be deemed inauthentic or authentic based on a person's present settled disposition.

Consider, for instance, a cognitively able person who has signed an advance directive rejecting an invasive treatment, who falls unconscious and comes to need it. In such cases, the decisionmaker has no reason for thinking that the preferences expressed in the advance directive have changed and, thus, no reason for assuming they are inauthentic. Withholding the rejected treatment option, then, does not violate the principle of determining authenticity.

Note, that this conclusion would hold even if the person was living with dementia, as long as there was no evidence of a change in disposition. Say, for instance, that this person has rejected a particular treatment option because of the likelihood that it will leave them doubly incontinent, which is something they regard as an intolerable indignity. It would be a mistake to assume that, just because they are living with dementia, this value set has disappeared. As such, it should be assumed that this remains the person's settled disposition unless there has been any indication since the progression of the dementia that they no longer view double incontinence in these terms (if, for instance, they have had sporadic episodes of incontinence that have not seemed to bother them).

This case illustrates an important component of the principle of determining authenticity, as I am arguing for it here. The principle does not require a decisionmaker to reject an advance directive wholesale, just because the person who wrote it has now developed

dementia. Rather, they should use all the information available to them to determine whether the preferences and values expressed in the advance directive remain authentic, and to consider the same of new expressions of values and preferences. Moreover, there remains a role for a kind of *reasonable person* standard, in the sense that we are being asked to consider what a reasonable person with the patient's settled disposition would choose. A non-dominating decision is one which has been reached by a decisionmaker who has engaged in the external reflective process and shown due regard to the determined authentic mental elements.

Of course, in clinical practice, there are decisions that must be made which are not as easy to solve as the case of harmlessly extending the life of a person living with dementia, or engaging in a procedure that might leave them incontinent. Often decisions must be made which reasonable people would disagree about. For instance, intensive cancer interventions which put the body under extreme stress and leave the person in constant pain and sickness may strike many as an undesirable trade for a few more years of life. Moreover, violent cardio-pulmonary resuscitation, which is highly likely to break ribs and, in the case of very elderly and frail people, unlikely to work may also not seem to be worth the trade for the chance of a few more years.<sup>53</sup> Despite this increased complexity, however, the principle of determining authenticity does offer a path to making a decision.

Imagine Margo has a DNCPR which stipulates that she did not wish to be resuscitated because she did not want to live with dementia any longer than was necessary. If this were the only stated reason for signing it, then we might be tempted to overrule it. However, suppose it also stipulates that she understood this intervention was unlikely to work if she was frail, so did not want to be harmed by it, but rather, she wanted to be allowed to slip away in peace. In this case, while the prior motivation can reasonably be judged inauthentic, there is little indication that her settled disposition regarding risk-taking and aversion to pain has changed. No reasonable person with Margo's present preference-set, even with the preference to avoid

living with dementia removed, would want to be subjected to this violent process for only a very low chance of survival. As such, adhering to the DNCPR would not violate the principle of determining authenticity.

Regarding an advance directive to refuse invasive cancer treatment, we again need to consider whether anything about Margo's settled disposition has changed regarding her stated reasons for opposing intervention. Once again, if her only reason for signing such a directive were to avoid living a life with dementia, then we have ample evidence that this is no longer an authentic preference. However, if the stated reasons include an aversion to harms the treatment involves and an unwillingness to take risks when success is uncertain, then there is no reason to suppose these preferences are now inauthentic, unless something in her present disposition has changed.

Cases, no doubt, can become even more complex than this. It is possible, even likely that different decisionmakers following the process I have set out here will come to different conclusions on the best treatment pathways when odds are unpredictable. This, however, is not a reason to abandon the account. The principle of determining authenticity does not require that a decisionmaker makes a perfect decision, especially not in situations of uncertainty. It would be overly demanding to require this because, after all, we do not always make perfect decisions when deciding for ourselves. Rather, it requires that the decisionmaker uses all of the evidence available to them to construct a preference-set that represents the person's authentic, settled disposition, in order to make a decision which a reasonable person with that preference-set would come to.

This process entails, of course, that advance directives ought to be defeasible, but it also says something about how they should be drawn up. It may not always be the case that people are asked to state their reasoning for ruling out particular interventions. If they do not do so, however, the principle of determining authenticity becomes much harder to follow,

because it becomes much more difficult to determine whether or not the values and preferences that led that person to their decision still hold as part of their settled disposition. As well as requiring that they are viewed as defeasible then, establishing the principle requires that advance care planning must involve a discussion of the reasons behind the decisions that have been made.

## **Conclusion**

The principle of determining authenticity, in order to do its guiding work, recommends establishing processes by which the authenticity of the preferences and values stated in an advance directive can be determined. In cases that do not feature dementia, the decisionmaker ought to begin from a presumption that nothing meaningful has changed since the signing of the advance directive and, as such, there is no reason to suspect it is inauthentic. However, in cases like Margo's, there will likely need to be a process of external reflection to determine whether or not her settled, present disposition indicates that she will repudiate all or part of her advance directive.

The principle of determining authenticity, as I have argued here, is intended to replace respect for autonomy and beneficence as the standard for substitute decision-making. To adhere to it, the decisionmaker must consider all evidence of values and preferences they have, including the advance directive and any more recent information that contradicts it, to construct a set of preferences that can be deemed authentic, in the sense of being non-alienating. Once this is established, they must make a decision which shows due regard to these preferences, based on what a reasonable person with those preferences would decide. Sometimes this may be imprecise, sometimes people may disagree but, as long as this process has been followed, the person in question will be free from domination.

- 
- <sup>1</sup> A. D. Firlik, "Margo's Logo," *JAMA: Journal of the American Medical Association* 265, no. 2 (1991): 201, at 201.
- <sup>2</sup> R. Dworkin, *Life's Dominion: An Argument About Abortion, Euthanasia, and Individual Freedom* (London: Vintage Books, 1994), at 226.
- <sup>3</sup> D. C. Thomasma and E. D. Pellegrino, "The Role of the Family and Physicians in Decisions for Incompetent Patients," *Theoretical Medicine* 8, no. 3 (1987): 283-286.
- <sup>4</sup> K. B. Hirschman, J. M. Kapo, and J. H. Karlawish, "Why Doesn't a Family Member of a Person With Advanced Dementia Use a Substituted Judgment When Making a Decision for That Person?," *American Journal of Geriatric Psychiatry* 14, no. 8 (2006): 659-661.
- <sup>5</sup> N. L. Cantor, "The Bane of Surrogate Decision-Making Defining The Best Interests of Never-Competent Persons," *Journal of Legal Medicine* 26, no. 2 (2005): 155-205.
- <sup>6</sup> D. W. Brock, "Death and Dying: Euthanasia and Sustaining Life: Ethical Issues," in *Encyclopedia of Bioethics*, ed. Warren T. Reich (New York: Simon and Schuster, 1995), 563-572.
- <sup>7</sup> Dworkin, *Life's Dominion*, 232.
- <sup>8</sup> N. Berlinger, "You Can't Always Get (or Give) What You Want: Preferences and Their Limits," *Hastings Center Report* 48, no. 3 (2018): 40; J. A. Wasserman and Mark. C. Navin, "Capacity for Preferences: Respecting Patients with Compromised Decision-Making," *Hastings Center Report* 48, no. 3 (2018): 31-39.
- <sup>9</sup> D. Brudney, "Choosing for Another: Beyond Autonomy and Best Interests," *Hastings Center Report* 39, no. 2 (2009): 31-37; D. Brudney and J. Lantos, "Agency and Authenticity: Which value grounds patient choice?," *Theoretical Medicine and Bioethics* 32, no. 4 (2011): 217-27).
- <sup>10</sup> Dworkin, *Life's Dominion*, 205.
- <sup>11</sup> Dworkin, *Life's Dominion*, 226.
- <sup>12</sup> Dworkin, *Life's Dominion*, 223.
- <sup>13</sup> Dworkin, *Life's Dominion*, 223.
- <sup>14</sup> Dworkin, *Life's Dominion*, 224.
- <sup>15</sup> Dworkin, *Life's Dominion*, 226.
- <sup>16</sup> Dworkin, *Life's Dominion*, 224.
- <sup>17</sup> Dworkin, *Life's Dominion*, 224.
- <sup>18</sup> Dworkin, *Life's Dominion*, 228.
- <sup>19</sup> Dworkin, *Life's Dominion*, 229.
- <sup>20</sup> Dworkin, *Life's Dominion*, 229.
- <sup>21</sup> Dworkin, *Life's Dominion*, 201.
- <sup>22</sup> Dworkin, *Life's Dominion*, 201.
- <sup>23</sup> Dworkin, *Life's Dominion*, 202.
- <sup>24</sup> Dworkin, *Life's Dominion*, 205-206.
- <sup>25</sup> Dworkin, *Life's Dominion*, 230.
- <sup>26</sup> Dworkin, *Life's Dominion*, 230.
- <sup>27</sup> Dworkin, *Life's Dominion*, 230.
- <sup>28</sup> Dworkin, *Life's Dominion*, 235.
- <sup>29</sup> A. Jaworska, "Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value," *Philosophy & Public Affairs* 28, no. 2 (1999): 105-138, at 109.
- <sup>30</sup> Jaworska, "Respecting the Margins of Agency", 119.
- <sup>31</sup> Jaworska, "Respecting the Margins of Agency", 116.
- <sup>32</sup> Jaworska, "Respecting the Margins of Agency", 113.
- <sup>33</sup> Jaworska, "Respecting the Margins of Agency", 125-126.
- <sup>34</sup> Jaworska, "Respecting the Margins of Agency", 119-124.
- <sup>35</sup> Jaworska, "Respecting the Margins of Agency", 119.
- <sup>36</sup> Dworkin, *Life's Dominion*, 225.
- <sup>37</sup> R. L. Berghmans, "Advance Directives for Non-Therapeutic Dementia Research: Some Ethical and Policy Considerations," *Journal of Medical Ethics* 24, no. 1 (1998): 32-37, at 36.
- <sup>38</sup> R. Dresser, "Dworkin on Dementia: Elegant Theory, Questionable Policy", *Hastings Center Report* 25, no.6, 32-38, at 36.
- <sup>39</sup> G. J. Westerhof, "Life Stories and Mental Health: The Role of Identification Processes in Theory and Interventions," *Narrative Works* 2, no. 1 (2012), 106-128.
- <sup>40</sup> J. M. G. Williams, "Capture and Rumination, Functional Avoidance, and Executive Control (CaRFAX): Three Processes That Underlie Overgeneral Memory," *Cognition and Emotion* 20, no. 3-4 (January 2007): 548-568.
- <sup>41</sup> P. Penn, "Rape Flashbacks: Constructing a New Narrative," *Family Process* 37, no. 3 (1998): 299-310.
- <sup>42</sup> Dworkin, *Life's Dominion*, 227-228.



- 
- <sup>43</sup> J. Christman, *The Politics of Persons: Individual Autonomy and Socio-historical Selves* (Cambridge: Cambridge University Press, 2009), 155.
- <sup>44</sup> Christman, *The Politics of Persons*, 152-153.
- <sup>45</sup> Christman, *The Politics of Persons*, 147.
- <sup>46</sup> Christman, *The Politics of Persons*, 145.
- <sup>47</sup> Christman, *The Politics of Persons*, 145.
- <sup>48</sup> L. F. Feinberg and C. J. Whitlatch, "Are Persons With Cognitive Impairment Able to State Consistent Choices?," *Gerontologist* 41, no. 3 (2001), 374-382.
- <sup>49</sup> Dresser, *Dworkin on Dementia*, 31-35.
- <sup>50</sup> E. Walsh, "Cognitive Transformation, Dementia, and the Moral Weight of Advance Directives," *The American Journal of Bioethics* 20, no. 8 (2020): 54-64, at 59.
- <sup>51</sup> P. Pettit, *On the People's Terms: A Republican Theory and Model of Democracy* (Cambridge: Cambridge University Press, 2012), 58.
- <sup>52</sup> Pettit, *On the People's Terms*, 59.
- <sup>53</sup> G. E. Applebaum, J. E. King, and T. E. Finucane, "The Outcome of CPR Initiated in Nursing Homes," *Journal of the American Geriatrics Society* 38, no. 3 (1990): 197-200.