# 6 Signed, sealed, but still 'me'? The conventionalist challenge to prospective autonomy

*Thomas Schirmer and Nils-Frederic Wagner*

## Introduction

When contemplating end-of-life care instructions, people frequently express sentiments like, 'I wouldn't be myself in that situation', or, 'I don't want that to be what's left of me'. Though such statements may initially seem reasonable, they imply significant normative and metaphysical challenges. How can I, as the person I am today, make decisions regarding the life and death of a future self whom I might not identify with—or who may no longer, in any meaningful sense, be 'me'? To address these challenges, we explore prominent arguments positioned at the intersection of philosophy of mind and bioethics.

The first field is the metaphysical discourse on the nature of personhood and personal identity, a discourse rooted in Enlightenment thought and launched by John Locke in the 17th century. Locke, in his influential *An Essay Concerning Human Understanding,* attempted to disentangle the concept of personal identity from contentious theological or metaphysical assumptions by analyzing the unity of a person as a relation inherent to consciousness and unattached to any kind of substance, like, for instance, an immortal soul. Nevertheless, the ensuing discourse focused on the analytical topos of how persons persist through time. After excluding the metaphysics of Cartesian dualism, physical continuity and psychological continuity were seen as the most plausible contenders, until, in the last two decades of the 20th century, an alternative perspective gained traction. The introduction of conventionalism to the subject of personal identity rapidly advanced the view that personhood and personal identity are, at least in part, socially constructed conventions, thereby challenging the previously dominant realist, 'non-conventionalist' perspectives.

The second discourse is bioethical and centers on the normative question of prospective autonomy in medical contexts. Bioethics flourished from the mid-20th century onward, prompted by the overwhelming progress of modern medicine and the accompanying imperative to safeguard human rights and dignity amid increasingly complex treatment decisions. How, for instance, can we secure a patient's fundamental right to autonomy in

DOI: 10.4324/9781003645122-9

scenarios involving unconsciousness, permanent vegetative states, or severe dementia? In the late 1960s, Chicago lawyer Luis Kutner addressed this issue by developing the concept of advance directives—legally binding documents that allow patients to maintain autonomy over their medical decisions even if they later become unable to make those decisions themselves (Benzenhöfer & Hack-Molitor, 2009). However, two decades later, Rebecca Dresser and John Robertson (Dresser, 1986; Dresser & Robertson, 1989) challenged the legitimacy of advance directives in cases of a profound psychological change and cognitive decline. They argued that conditions like Alzheimer's disease can alter a person so drastically that it calls into question whether the individual at the onset of the illness remains the same person in its later stages. Do patients with dementia, over the course of their illness, effectively become 'someone else' (DeGrazia, 1999)? And if so, what implications does this have for the validity of their advance directives? By the early 21st century, the treatment of patients with dementia-related illnesses had become both a pressing real-world medical challenge and a key test case for theories of personal identity and autonomy in medical decision-making.

After briefly introducing conventionalism in personal identity (Section 'Conventionalism about personal identity') and providing a thorough analysis of the 'someone else problem' in severe dementia cases (Section 'Advance directives and the "someone else problem"'), we demonstrate in Section 'Why care for conventional persons?' how prominent theories addressing the 'dementia dilemma' implicitly rely on conventionalist views—whether or not their authors acknowledge, explicitly endorse, or even oppose such reasoning. As a result, we argue that it is conceptually inconsistent to uphold both a non-conventionalist view of personal identity and the normative authority of advance directives. Conventionalist reasoning appears to be an essential foundation for the institutional framework of advance directives, which, in turn, could be seen as an argument in favor of conventionalism itself. Suggesting that advance directives may play a role in shaping our understanding of personhood. Alternatively, proponents of a non-conventionalist stance might reconsider whether, in some cases, the significance of advance directives lies more in preserving the dignity of their signatories than in benefiting their recipients. Our analysis offers new insights into whether and why, in situations of potential disruptions of personal identity, advance directives should be regarded as normatively binding. From a broader analytical perspective on personal identity, we explore how well orthodox, non-conventionalist theories hold up when applied to bioethical dilemmas of real-world significance.

## Conventionalism about personal identity

In the analytical discourse on personal identity, the term conventionalism is best understood by what it rejects: namely, the idea that personhood and personal identity are natural kinds—objective, metaphysical facts that exist

independently of human interpretation. Instead, conventionalism holds that the necessary and sufficient conditions for an individual's survival—and even the question of whether they remain a person over time—are, at least in part, shaped by social conventions. This view introduces a striking conundrum.

On the one hand, the concept of 'person' plays a prescriptive role in normative and legal discussions, where attributing personhood to an individual confers a unique moral status. Various theories propose different capacities—often human-specific—as the basis for this status, suggesting that determining who deserves the rights and recognition of personhood depends on identifying a threshold of relevant qualities. This pursuit, which seeks to justify our shared intuitions about whom we should or should not treat as a person, appears to be deeply influenced by social conventions. Let us refer to this as the *personhood question*. Conversely, we also use the concept of 'person' descriptively in normative and legal contexts: Is the person before me responsible for actions they committed in the past? Will they be entitled to the future benefits of their actions today? It seems counterintuitive to suggest that personal identity over time depends on social conventions about what defines personhood. After all, this would imply that my very existence as a person now is contingent on socially constructed criteria (Merricks, 2001). Instead, the notion of personal identity appears to point to an underlying metaphysical essence—grounded in certain capacities of personhood—that sustains identity over time. Let us refer to this as the *persistence question*.

Three possible responses to this challenge emerge:

A The non-conventionalist solution: A non-conventionalist might argue that applying conventionalism to normative and legal issues commits a normative fallacy. If we could determine what fundamentally constitutes personhood and numerical identity over time—whether through physical continuity, psychological continuity, or another criterion—we could resolve normative questions based on that foundation, with our intuitions naturally aligning. This approach is pursued by authors like McMahan and Schechtmann (discussed in Sections 'McMahan: Degrees of egoistic concern' and 'Schechtman: Why accord a place in "person-space"').
B The conventionalist solution: A conventionalist, by contrast, would argue that personhood itself is a convention-based concept. Challenging the orthodox realist perspective, they maintain that both personhood and personal identity over time are fundamentally shaped by social conventions. This is the view implicitly endorsed by authors such as Buchanan and Brook and Kuczewski, even if they do not explicitly label it as conventionalism.
C The hybrid solution: A third perspective holds that one can adopt a conventionalist stance on the *personhood question* while remaining a non-conventionalist about the *persistence question*. For instance, animalist positions often align with this view, asserting that we persist over time not as persons but as biological organisms. However, this

raises the challenge of explaining how such a conception of numerical identity meaningfully connects to normative concerns (see Luzio, 2025). A version of this approach is advanced by DeGrazia (discussed in Section 'DeGrazia: Prospective narrative identity as persistence conventionalism').

Braddon-Mitchell and Miller (2004) support (B) by contending that the very existence of a human person is indeed dependent on conventions. These conventions may go unnoticed in everyday life precisely because they are so well-established; they operate as settled conventions that we take for granted. Nonetheless, according to Braddon-Mitchell and Miller, these conventions are fundamental in defining personal identity over time:

> For instance, it is a settled convention that persons have some attitude of self-concern toward physical and psychological continuants. Equally, it is pretty much settled that persons have no attitude of self-concern towards entities in the future that are neither physical nor psychological continuants.
>
> (Braddon-Mitchel & Miller, 2004, p. 463)

Braddon-Mitchell and Miller argue that classic puzzle cases in the discourse on personal identity seem difficult because psychological and physical continuity can diverge. While our established conventions are generally sufficient, they become inadequate in such cases. Although Braddon-Mitchell and Miller primarily illustrate their point using thought experiments involving teletransportation or brain transplants, they acknowledge that similar divergences occur in real life: 'We sometimes speak as though in reality these do not come apart, but of course sometimes they do, as in cases of amnesia, infancy, and dementia' (Braddon-Mitchel & Miller, 2004, p. 472, footnote 26). Thus, despite the analytical discourse on the metaphysics of personal identity—often centered on counterfactual thought experiments—there exists an equally significant practical discourse within bioethics (Wagner, 2022). This discussion, grounded in everyday medical care, explores identity and autonomy in real-world contexts, particularly regarding the legitimacy of advance directives. A central issue in this debate is the 'someone else problem'—the question of whether a person who has undergone a profound loss of rational capacity and psychological change remains the same person who originally issued the directive.

## Advance directives and the 'someone else problem'

In the latter half of the 20th century, a societal norm emerged affirming that any self-determining individual (*S*) has the fundamental right to make autonomous medical decisions concerning themselves at a given time (*t*)—a

principle known as *informed consent*. This right is now widely understood to extend to cases where:

a Decisions made at time ($t$) will only be enacted at a future time ($t^*$), and
b The future self ($y$), who will be affected at ($t^*$), may no longer have the capacity for competent self-determination in a medical sense.

Macedo et al. (2023) observe that since the late 1960s, the concept of *prospective autonomy* has developed through the introduction of legally binding advance directives in many Western legal systems. These directives take two primary forms:

1 *Durable Power of Attorney for Healthcare*: The present self ($S$) legally appoints a proxy to make medical decisions on behalf of the future self ($y$), should $S$ lose the capacity for informed consent at time ($t^*$).
2 *Living Will:* The present self ($S$) provides binding instructions to medical professionals regarding the acceptance or refusal of specific medical treatments for the future self ($y$) at time ($t^*$). While these instructions are legally enforceable, certain forms of basic care are typically excluded from refusal.

These two forms of advance directives serve as complementary safeguards for prospective autonomy—one by designating a trusted decision-maker, the other by pre-specifying treatment preferences. Creating a living will can be seen as a logical extension of informed consent, reinforcing the prominent ethical principle of respect for autonomy (Beauchamp & Childress, 2019). However, two key objections challenge the assumption that advance directives are simply an extension of informed consent.

The first objection is normative: the apparent temporal extension from informed consent to a living will does more than merely carrying a decision across time—it fundamentally alters the nature of the decision itself. What might be a straightforward cost-benefit analysis in a present medical decision becomes, in a living will, a deeply value-laden judgment about the worth or desirability of one's future life. This shift raises a critical question: is the relatively low threshold for autonomous medical decision-making—based on Beauchamp and Childress's 'normal chooser' model—adequate for such a profound and far-reaching determination?

Beauchamp and Childress reject theories of autonomy that require higher-order reflection as a prerequisite for medical decision-making. Instead, they explicitly define autonomy 'as a way for ordinary persons to qualify as deserving respect for their autonomous choices even when they have not reflected on their preferences at a higher level' (Beauchamp & Childress, 2019, p. 101). However, the concept of the living will introduces a deeper metaphysical challenge: in standard cases of informed consent, the implicit assumption is that the self (the rational human person) makes a decision

about medical treatment for itself (the biological human organism). In everyday medical decision-making, this dual perspective poses little friction—one can naturally conceive of oneself as both a psychologically continuous person and a living human being. However, living wills, which anticipate future scenarios, introduce a troubling question: What happens if these two dimensions of selfhood—the psychological continuer (the person) and the biological continuer (the human organism)—diverge? If I, as a psychological entity, and I, as a biological being, do not align in the future, can the original *S* still be considered sufficiently continuous with the future *y* to justify the binding authority of the living will? As DeGrazia observes, 'advance directives are supposed to give guidance for one's own medical care' (DeGrazia, 1999, p. 378). This raises a fundamental issue: how much and what type of change can *S* undergo while still being considered identical to *y*?

Originally articulated by Rebecca Dresser in the 1980s (Dresser, 1986; Dresser & Robertson, 1989), this problem has become central to debates on personal identity and autonomy in medical decision-making (e.g., Buchanan & Brock, 1990; DeGrazia, 2005; McMahan, 2002). The issue, often referred to as the 'dementia dilemma', is typically framed as follows:

1 *S* is a rational human person, capable of self-determination, who is diagnosed with dementia at time *t*. *S* has long valued autonomy and intellectual independence.
2 S assumes personal identity over time with a future self, *y*, at time *t** (a later stage of dementia).
3 After thorough medical counsel, *S* creates an advance directive at time *t*, stating that *y* should not receive life-sustaining treatment once they lose self-determination at time *t**.
4 At time *t**, *y* has indeed lost self-determination but appears content and expresses a natural will to live.
5 At *t**, a life-threatening situation arises that requires medical intervention—one explicitly refused in *S*'s advance directive. Without treatment, *y* will die.

The central question in the dementia dilemma is whether *S*'s advance directive can legitimately exert authority over *y* in this scenario. Within the traditional bioethical discourse on advance directives, premise (2)—the assumption of personal identity over time between *S* and *y*—is largely taken for granted. From this perspective, the dementia dilemma is framed in terms of how to appropriately reconstruct autonomy and weigh the normative relationship between two distinct forms of agency: How should one balance the autonomous decision of S against the present, non-autonomous will of y? (Jaworska, 2017). However, from a metaphysical standpoint, premise (2) requires closer scrutiny. Is *S* truly identical to *y* in a way that justifies imposing *S*'s past wishes on *y*'s present reality? To answer this question, we must first analyze the nature of *S*'s transformation throughout the progression of dementia. A meaningful exploration of

the ethical and metaphysical issues surrounding advance directives in dementia cases hinges on distinguishing four key concepts:

1 Decision-making capacity (DC): The ability to understand, deliberate, and make autonomous medical decisions.
2 Psychological continuity (C): The preservation of memory, personality traits, values, and a coherent sense of self over time.
3 Personal identity over time (I): The philosophical question of whether *S* and *y* are truly the same entity across time.
4 Personhood (P): The status of being a moral or legal person, often linked to specific cognitive capacities.

While these dimensions frequently overlap in practice, they are distinct in principle. Conflating them risks obscuring the core ethical and metaphysical questions surrounding the normative authority of advance directives. The next step is to examine each in detail.

Loss of decision-making capacity (DC): Losing decision-making capacity means *y* can no longer make informed, autonomous choices. In medical ethics, this often triggers substituted judgment, relying on prior wishes or appointed surrogates. However, DC does not necessarily imply the loss of psychological continuity (C), personal identity over time (I), or personhood (P); it reflects a functional impairment rather than a complete transformation of the self.

Loss of psychological continuity (C): Psychological continuity refers to the persistence of key mental states—memories, beliefs, desires, and personality traits—over time. Unlike psychological *connectedness*, which requires only partial mental links, continuity demands a sufficient degree of such connections. Given *S*'s cognitive decline, one might argue that *y* lacks the required psychological connectedness to maintain (C), challenging the idea that *y* is still *S* in a personal sense. However, (C) does not necessarily entail the loss of (I) since some theories maintain that physical continuity alone preserves numerical identity.

Loss of personal identity over time (I): personal identity in the numerical sense means that *y* at $t^*$ is strictly the same entity as *S* at *t*. Some theories, particularly those prioritizing physical continuity, hold that *y* remains numerically identical to *S* as long as they share the same living biological organism. However, if psychological continuity is deemed essential to identity, then (C) necessarily implies (I), meaning *y* and *S* are no longer numerically identical.

Loss of personhood (P): Personhood is typically defined by autonomy, self-reflection, and moral agency. If *y* lacks decision-making capacity and suffers severe psychological impairment, some may argue that *y* is no longer a person in an ontological sense (P), as core attributes of personhood are absent. This classification carries profound ethical implications, as it affects moral status and rights. While (P) does not require (I), it necessarily entails (C) since psychological connectedness alone cannot sustain psychological continuity.

The implications for the dementia dilemma: For *S*'s living will to take effect, *S* must lose decision-making capacity (DC). However, the extent of *S*'s psychological transformation remains unclear. This gives rise to three key claims:

a (DC) = (C): Losing decision-making capacity (DC) necessarily entails the loss of psychological continuity (C), meaning *y* at *t** is no longer the same person as *S* at *t*.
b (DC) = (P): Losing decision-making capacity (DC) necessarily entails the loss of personhood (P), meaning *y* at *t** is no longer a person *at all*.
c (DC) = (I): Losing psychological continuity (C) necessarily entails the loss of personal identity (I), meaning *y* at *t** is not numerically identical to *S* at *t*.

As Jaworska emphasizes, advance directives are

> best suited for the contexts for which they were first developed in the law—conditions involving loss of consciousness such as persistent vegetative state—where the patient in the current incompetent state cannot have interests potentially different from the interests of the person he used to be.
>
> (Jaworska, 2017)

In cases of persistent vegetative state (PVS), the underlying condition typically results in the simultaneous loss of decision-making capacity (DC), psychological continuity (C), and personhood (P), thereby affirming claims (a) and (b). For proponents of physical continuity as the basis of personal identity over time, the normative authority of an advance directive in such cases follows from rejecting claim (c), maintaining that $S = y$. Conversely, those who consider psychological continuity essential to personal identity accept even claim (c), concluding that $S \neq y$. From this perspective, *S*'s surviving interests justify the authority of the advance directive in terms of quasi-property rights over the mere biological shell of *y* (see Buchanan & Brook, 1990, discussed in Section 'Buchanan and Brook: The threshold of psychological continuity as convention).

At first glance, both perspectives seem to affirm *S*'s authority over *y*, as the questionable equation (DC) = (P) in claim (b) allows them to sidestep deeper metaphysical concerns. However, a greater challenge arises when loss of self-determination occurs under less drastic yet permanent conditions, where the incompetent patient develops seemingly powerful new interests in their altered state (Jaworska, 2017). The dementia dilemma exemplifies this complexity, revealing that the relationship between personal identity and advance directives is far more intricate than previously assumed.

In cases of dementia, claim (b) encounters significant complications. While self-determination is widely considered an essential aspect of personal identity, it remains unclear whether this capacity is merely sufficient or genuinely

necessary for personhood. As dementia progresses, situations arise where *y*, though lacking decision-making capacity (DC), still seems to qualify as a person (P). To resolve this tension, one could either reject the intuition that *y* remains a person—a conclusion Marya Schechtman challenges:

> If we really thought that the person was gone in the strictest sense of the term, it would be far less painful to see 'her' in this condition than it often is for people to see their loved ones suffering from dementia.
>
> (Schechtman, 2010, p. 276)

Alternatively, we could reject claim (b) and affirm that (DC) does not necessarily entail (P). This would imply a distinction within the category of persons: those with (DC), and those without. If (DC) is crucial for psychological (and thus personal) continuity between *S* and *y*, then the loss of (DC) might mean that *y* at *t** is, in effect, a new person—someone else.

The 'someone else problem' arises most clearly in theories rooted in *person essentialism* (DeGrazia, 1999), which define personhood as the core of human identity and treat psychological continuity as the primary criterion for personal identity over time. These theories effectively equate psychological continuity (C) with personal identity (I), thus upholding claim (c). As DeGrazia states: 'No human being who is a person can ever exist as a non-person; personhood is a necessary condition for the existence of a human being who is ever a person' (DeGrazia, 1999, p. 380). If (DC) disrupts (C), it often suffices to sever personal identity, even if it does not fully eliminate personhood (P). Alternatively, (DC) may at least coincide with psychological changes significant enough to break (C) in a comparable way. In either case, if *S* and *y* are not numerically identical, then *S*'s past preferences would lack authority over *y*'s present treatment: 'The former self's preferences would have no particular authority to govern the incompetent patient's treatment' (Dresser & Robertson, 1989, p. 236).

The dementia dilemma appears less troubling for theories that reject *person essentialism* and instead ground personal identity over time in physical continuity, distinguishing the loss of psychological continuity from the loss of personal identity over time. Within this framework, despite psychological discontinuity between *S* and *y*, *S* at *t* and *y* at *t** remain the same biological human being. However, this view raises critical questions: How can *I*, as the human animal *S*, conceive of myself primarily as an organism, given that rationality is neither necessary nor sufficient for being human? More importantly, how could *I*, as the biological entity *S*, care about the future entity *y* in the deeply personal way I actually do? In other words, mere numerical identity over time does not necessarily entail the *egoistic concern* (McMahan, 2002, p. 41; see also Section 'McMahan: Degrees of egoistic concern') that *S* typically has for *y*. As David Shoemaker aptly asks: 'How could the continuity of meat serve to explain a relationship of minds?' (Shoemaker, 2016, p. 305). Even within a framework emphasizing physical continuity,

psychological continuity arguably remains the primary foundation of egoistic concern—and thus of the legitimacy of advance directives. Consequently, the loss of psychological continuity supports viewing the surviving entity *y*, despite personal identity over time, as effectively someone else, raising legitimate concerns about the authority of an advance directive issued by the former person (DeGrazia, 1999).

Despite DeGrazia's legitimate worries, proponents of both psychological and physical continuity generally do not reject the authority of advance directives in the dementia dilemma. This raises a key question: How can one challenge psychological continuity, while still maintaining that *S*'s past preferences should meaningfully influence y's treatment? As we will explore, many of the most prominent theories of personal identity—when addressing the 'someone else problem'—appear to rely, whether explicitly, implicitly, or even while denying it, on conventionalism in navigating the practical complexities of the dementia dilemma.

## Why care for conventional persons?

### *Buchanan and Brook: The threshold of psychological continuity as convention*

Building on their assumption that psychological continuity is essential for personal identity over time, Buchanan and Brook (1990) examine the 'someone else problem' from a person-essentialist perspective. They present the issue in its most fundamental form: if psychological continuity between *S* and *y* is disrupted, yet *y* remains a person, then *y* is, in effect, 'someone else'. As we will see, their proposed solution implicitly relies on a conventionalist view of personhood.

To clarify their argument, Buchanan and Brook outline two distinct cases within the dementia dilemma. In the first scenario, *y* at *t** lacks psychological continuity with *S* and is no longer a person at all. In this case, the authority of an advance directive is justified by what the authors call *S*'s 'surviving interests'. They argue that persons, through a performative 'act of will' (1990, p. 116), can generate interests that persist beyond their loss of decision-making capacity or even beyond their death. Since these interests are tied to specific intended outcomes or commitments, they continue independently of the person's ongoing identity. By drafting an advance directive, *S* effectively establishes quasi-property rights over the future non-person *y*. However, Buchanan and Brook acknowledge that rights based on surviving interests do not carry the same normative weight as those grounded in self-determination (1990, p. 116). Consequently, an advance directive in this case may be more easily overridden than a contemporaneous informed consent decision, particularly when significant harm or benefit is at stake. The second case, however, presents a greater challenge. Here, *y* at *t** remains a person but lacks sufficient psychological continuity with *S*, thus creating a

true 'someone else' scenario. In this situation, Buchanan and Brook argue that *S*'s advance directive cannot hold absolute normative authority over *y*. Their conclusion is striking:

> since the value of preserving some of our most basic institutions and practices speaks in favor of setting the threshold of psychological continuity necessary for personal identity low, and since there is nothing of comparable weight on the other side of that balance, we clearly ought to set (or rather leave) the threshold low.
>
> (Buchanan & Brook 1990, p. 175)

By setting the necessary threshold for psychological continuity as low as possible, Buchanan and Brook seek to minimize the number of cases where *S* and *y* are no longer meaningfully connected, yet *y* remains a person. For this limited subset of cases, they argue that as the psychological connectedness between *S* and *y* progressively deteriorates, the normative authority of the advance directive should diminish in proportion. Crucially, Buchanan and Brook contend that the threshold for psychological continuity should be determined pragmatically, in accordance with societal norms—a position that aligns explicitly with a conventionalist view of personal identity. As established in Section 'Conventionalism about personal identity', those who regard personhood as essential to personal identity over time and simultaneously argue for a low threshold of psychological continuity in cases of cognitive decline are, in effect, adopting a conventionalist stance on both the *persistence* and the *personhood question*. In other words, if we maintain that a human being persists as the same person over time while also setting a minimal threshold for psychological continuity, then the criteria for personhood itself must be correspondingly low from the outset.

This perspective suggests that personhood is not contingent upon a high degree of cognitive or psychological function; rather, it can endure as long as some minimal form of continuity remains. By grounding the conditions for psychological continuity within prevailing societal norms, Buchanan and Brook ultimately position their argument within a conventionalist framework of personhood.

#### *Kuczewski: Community-based survival of self-interests*

Mark Kuczewski builds on Buchanan and Brook's reconstruction of the 'someone else problem' and their proposed solution but critiques their 'individualistic interpretation of the self', arguing that it makes their theory susceptible to common challenges faced by psychological continuity theories (Kuczewski, 1994, p. 36). He takes issue with their notion of 'surviving interests'—the idea that a person can retain quasi-property rights over their non-person successor—contending that it relies on the mistaken assumption of a persisting 'I' that can possess such interests.

In contrast, Kuczewski locates these interests not within an enduring self but within the collective memories of the community, both personal and institutional (Kuczewski, 1994, p. 42). He advocates for a more fluid, communitarian conception of the self (Kuczewski, 1994, p. 36), which, in his view, better reflects how society navigates the complexities of personhood and personal identity—'phenomena with many components' (Kuczewski, 1994, p. 44). His argument that 'body and psychological continuity are intimately bound up with personal identity but not definitive of it in any straightforward manner' (Kuczewski, 1994, p. 44) anticipates Braddon-Mitchell and Miller's discussion of 'puzzle cases' in Section 'Conventionalism about personal identity'. Ultimately, Kuczewski concludes that in cases involving conscious but incompetent patients, determining whether the individual remains the 'same person' is not merely a descriptive question but a societal decision informed by descriptive elements (Kuczewski, 1994, p. 46, footnote 60). More explicitly than Buchanan and Brook, he embraces what we would now call a conventionalist approach. His communitarian view—that the unity of a personal life is granted by the community—finds a more refined articulation in Marya Schechtman's 'Person Life View', which we explore below.

### *McMahan: Degrees of egoistic concern*

Jeff McMahan (2002), in presenting his 'embodied mind' account, takes dementia as a paradigmatic case for exploring the relationships between psychological continuity, personal identity over time, and egoistic concern—the unique concern I have for my future self, which differs in both kind and degree from the concern I have for others. At the same time, he reinforces the normative authority of advance directives in severe dementia cases. Like Buchanan and Brook and Kuczewski, McMahan holds that psychological continuity is a necessary condition for personal identity over time. However, what distinguishes him from these authors is his rejection of an all-or-nothing approach to psychological continuity; instead, he argues that psychological continuity exists on a spectrum (McMahan, 2002, p. 40–42). While McMahan maintains that strong psychological continuity is required for personhood, he asserts that even weak, minimal continuity is sufficient for personal identity over time and egoistic concern. Following Parfit (1984), he emphasizes that psychological continuity—rather than identity per se—is what grounds egoistic concern. Accordingly, he differentiates between a 'psychological account of identity' and a 'psychological account of egoistic concern' (McMahan, 2002, p. 46–48). For McMahan, the degree of psychological unity within a life is determined by the physical, functional, and organizational continuity of brain regions associated with consciousness (McMahan, 2002, p. 79). As long as these brain regions retain minimal functionality, psychological continuity and personal identity over time persist. This leads him to conclude that we are not essentially persons, but rather 'embodied minds'.

Applying this theory to late-stage dementia, McMahan argues that even if personhood is lost, personal identity over time remains intact, thereby confirming premise (2) of the dementia dilemma. Since this view does not generate an obvious 'someone else problem' from a metaphysical standpoint, the authority of *S*'s advance directive should, in principle, be preserved. However, because psychological unity varies in strength depending on the extent of continuity in the conscious brain regions, the degree of egoistic concern that it is rational for *S* to have about *y* fluctuates accordingly (McMahan, 2002, p. 79). The psychological unity between '*S* today' and '*S* tomorrow' is greater than that between '*S* tomorrow' and '*S* in five years', making it rational for *S* to prioritize short-term self-interest over long-term self-states. This introduces what McMahan calls 'time-related interests' (TRI). In cases of severe dementia, he contends that although personal identity over time remains intact, the dramatic weakening of psychological unity causes *y*'s life to become increasingly alien from *S*'s. Beyond a certain threshold, *y* develops an independent existence, and their life should no longer be understood as a direct continuation of *S*'s. In this sense, while metaphysically identical, *S* and *y* become distinct in terms of egoistic concern—meaning that, in practical terms, *y* becomes 'someone else'. Since *y*'s current preferences, shaped by dementia, may be 'notoriously arbitrary, whimsical, and ephemeral' (McMahan, 2002, p. 498), McMahan frames the dilemma of honoring *S*'s advance directive as a conflict between *S*'s past autonomous preferences and *y*'s present best interests. Unlike Dworkin, who assumes that fulfilling *S*'s wishes inherently benefits *y*, McMahan argues that when psychological unity is weak, the best interests of *y* should be assessed more in terms of *y*'s present state than as part of *S*'s former life narrative (McMahan, 2002, p. 500). Given the choice between preserving *S*'s coherent life story and allowing *y* to continue an otherwise pleasurable existence, McMahan, somewhat surprisingly, favors *S*'s interests. He justifies this on the grounds that *y*'s time-related interest in continued life is weak—comparable, he suggests, to that of an animal (McMahan, 2002, p. 503).

McMahan's argument should be understood as conventionalist, as it relies on socially constructed conceptions of personhood, identity, and what constitutes a 'distortion' of one's life, rather than on any intrinsic or essential properties of selfhood. Two key conventionalist assumptions underpin his reasoning:

1 The Problem of Surviving Interests: If *S*'s egoistic concern no longer extends to *y* (because *y* is effectively 'someone else'), then it is questionable whether *y* is the appropriate bearer of *S*'s 'surviving interests'. Echoing Kuczewski's critique of Buchanan and Brook, McMahan's argument suggests that *S*'s surviving interests primarily persist in the minds of those around *y*—family, friends, and caregivers—rather than in *y* themselves. Since *y* lacks the capacity for self-reflection, it is the image held by others that influences *S*'s advance directive in the first place. In other words, it is

not some transcendent successor of *S* who determines the preservation of *S*'s identity, but rather the social re-identification of *y* with *S*. Thus, McMahan's defense of *S*'s life narrative implicitly depends on the same communitarian notion of surviving interests that Kuczewski openly endorses.

2 Socially Shaped Attitudes Toward Dementia: McMahan himself acknowledges that perceptions of dementia are shaped by societal norms, asking whether one would view the dementia dilemma differently 'if a period of dementia were universal after a certain age and were generally pleasant and contented the way childhood normally is' (McMahan, 2002, p. 499). This suggests that the transformation of *S* into *y* is not necessarily an essential or metaphysical problem, but one that hinges on cultural biases and the stigma surrounding cognitive decline. If dementia were perceived more neutrally, neither *S* nor those around *y* might view *y*'s existence as a distortion or an interruption of *S*'s life. McMahan's division of *S* and *y*'s existence into a 'healthy' and a 'demented' phase seems arbitrary unless understood conventionally: it is not the gradual weakening of psychological unity itself that justifies this division, but rather the social meaning attached to a diagnosis of dementia.

Thus, while McMahan's 'embodied mind' account may not be conventionalist in itself, his argument for the authority of advance directives certainly is.

***DeGrazia: Prospective narrative identity as persistence conventionalism***

David DeGrazia (2005) revisits his reformulation of the 'someone else problem' from a non-person-essentialist perspective. Grounded in the premise that physical continuity underpins personal identity over time, he defends the view that 'we' are essentially animals. However, drawing on Marya Schechtman's Self-Constitution View, he argues that while these animals possess the property of being persons, they construct self-narratives—an act that defines them as persons (DeGrazia, 2005, p. 81). As persons, we desire our self-narratives to continue unfolding, integrating future actions and experiences in a way that maintains psychological continuity between our present and future selves (DeGrazia, 2005, p. 81).

Emphasizing that 'disruption of narrative identity is normatively important despite numerical identity' (DeGrazia, 2005, p. 176), DeGrazia proposes a model of 'prospective psychological continuity' that links *S* to *y* and considers it sufficient for practical concern. If *S* identifies strongly enough with their future self *y*, this identification forms what he terms a 'weak narrative identity' (DeGrazia, 2005, p. 180). Under this framework, the conditions for the dementia dilemma are as follows: (C v ¬C; I; P v ¬P). If *S*, in exercising prospective autonomy, deliberately prioritizes their lifetime interests—including their future state as *y*—over *y*'s immediate welfare at *t**, then the authority of their advance directive can be justified. In this view, *S*'s autonomy rights, combined with their narrative-driven lifetime interests, support the directive's

authority over *y*'s present well-being: 'In effect, the autonomous agent determines how particular possible changes, such as dementia, affect their relationship with themselves over time' (DeGrazia, 2005, p. 196).

As Jaworska rightly observes, DeGrazia's argument allows the rational level of prudential concern for the future to be, at least in part, a choice of the earlier self. She critiques this position, arguing that 'prudential concern is a requirement of rationality and should not be a matter of choice' (Jaworska, 2017). The key underlying issue here is the implicit conventionalism in DeGrazia's argument. Consider an individual *S* who, as a philosopher, holds that high psychological continuity is essential for personal identity over time. Such an individual would likely identify less strongly with their future, demented self than DeGrazia, who prioritizes physical continuity. Consequently, the authority of an advance directive would vary depending on one's conception of personal persistence. Yet, while both philosophers would survive as animals, the idea that the legitimacy of their advance directives depends on their personal beliefs about personal identity over time suggests an underlying assumption: numerical identity between *S* at *t* and *y* at *t** is, at least in part, contingent on *S*'s own perspective on the *persistence question*. This leads to what we might call 'persistence conventionalism'—the notion that the continuity of personal identity is, to some extent, a function of one's beliefs about survival.

#### *Schechtman: Why accord a place in 'person-space'?*

In 'Personhood and the Practical' and *Staying Alive*, Marya Schechtman addresses the 'someone else problem' of the dementia dilemma by introducing her Person Life View (PLV) as a potential solution. After examining both psychological continuity theories (which she terms 'neo-Lockean') and physical continuity theories (such as Animalism), she finds both approaches limited by their reliance on a narrowly Lockean conception of personhood (Schechtman, 2010, p. 272). Traditional Lockean views, she argues, create a rigid dichotomy between 'person-making capacities'—such as moral agency, reflective self-consciousness, and reason—and 'animal features' shared with non-human animals, reducing the latter to 'merely biological factors' (Schechtman, 2010, p. 277). In practice, Schechtman contends, this classic framework often fails to bridge metaphysical and practical concerns, particularly in complex cases such as the dementia dilemma. To overcome these limitations, Schechtman proposes a broader understanding of personhood through the Person Life View, stating: 'A person is the being who lives a person life, and a single person continues for the duration of a single person life' (Schechtman, 2010, p. 278). Crucially, a *person life* is defined not merely by the accumulation of Lockean capacities but by participation in a culturally embedded, human-typical life. Schechtman rejects the notion that developing personhood is akin to layering rationality and self-awareness atop an animal substrate. Instead, she emphasizes that humans live animal lives infused with

human-specific cognitive and social functions, giving those lives their distinctive character (Schechtman, 2010, p. 279).

To illustrate, she contrasts the act of eating in everyday contexts with its transformation into a socially meaningful interaction in a wedding ceremony, where 'a glossy cake [is] ritualistically exchanged by bride and groom' (Schechtman, 2010, p. 279). Personhood, in this view, emerges from both the standard developmental trajectory of rationality and agency and from participation in culturally meaningful practices. For Schechtman, to *be* a person is to exist within *person-space*—a social and cultural framework that shapes and sustains human personhood. While the specifics of this framework vary across cultures, she argues that 'when we encounter other humans, we automatically see them as persons and interact with them as such' (Schechtman, 2014, p. 113).

This inclusive framework, Schechtman contends, accommodates individuals who diverge from typical developmental trajectories. Drawing on Hilde Lindemann's case of her hydrocephalic sister Carla, she highlights how individuals who lack full Lockean capacities can nonetheless inhabit *person-space*—provided they are recognized and integrated into social life (Schechtman, 2010, p. 280). Applying the PLV to the *someone else* problem in the dementia dilemma, Schechtman maintains that *S* and *y* remain the *same* person, thus affirming the normative authority of *S*'s advance directive. However, her framework also prompts a critical examination of the cultural biases and stigma inherent in *S*'s decision to draft the directive in its current form.

Despite its strengths, Schechtman acknowledges a potential weakness: if personhood under the PLV depends largely on how others perceive and position an individual, then the view may appear 'objectionably conventionalist' (Schechtman, 2010, p. 280). To counter this concern, she presents a two-step argument. First, she argues that personhood decisions should not be made on a purely case-by-case basis. Instead, she asserts that nearly all humans typically develop Lockean capacities and form human cultures, which, in turn, reliably grant *person-space* to humans—not simply based on rationality or self-consciousness, but through fundamental social interactions shared by all human communities. Even those who lack full Lockean capacities remain embedded in these networks of interaction: 'As we saw in Carla's case, "caring for" and "playing with" are also part of our interpersonal repertoire, as are "laughing together", "sharing a meal", "dancing with", etc.' (Schechtman, 2010, p. 281).

Thus, for Schechtman, personhood recognition is neither arbitrary nor conventional. It is, she argues, an automatic response (Schechtman, 2014, p. 114), an institutionalized aspect of human social structures, and ultimately rooted in our evolutionary history (Schechtman, 2014, p. 113). Because this person-recognition mechanism is universally present across human cultures, she concludes that Carla's inclusion in *person-space* is not a matter of mere convention. However, this argument raises concerns about circularity. Schechtman seems to claim that *because* we instinctively recognize

humans as persons, we *ought* to continue doing so. But is this an empirical fact or a normative prescription? To solidify her position, she must demonstrate that recognizing Carla as a person is *not only* a widespread human practice but also that it would be *rationally inconceivable* to do otherwise (a strong claim). Alternatively, she could aim for a weaker claim: that there are no real-world instances of humans being consistently excluded from personhood, nor cases where non-humans are consistently granted it. If she could successfully establish this weaker claim, it might lend indirect support to the stronger one.

To defend the PLV against concerns of being either over-inclusive or under-inclusive, Schechtman considers potential counterexamples. Regarding over-inclusivity, she argues that it would be *rationally untenable* to grant a beloved pet the same status as a human child. She illustrates this with a contrast between two scenarios: a family learns their human infant will never be able to talk, dress, or feed herself. Contrast this with a family being told their beloved poodle puppy will never develop these abilities (Schechtman, 2014, p. 121). While Schechtman concedes that non-human beings could theoretically attain personhood (e.g., a 'mutant poodle' with self-consciousness and language), she points out that, in practice, only human animals are granted *person-space* (Schechtman, 2010, p. 281). Regarding under-inclusivity, she highlights that *recognizing* someone as a person does not necessarily entail treating them well. She points to Antebellum slavery laws, which *presupposed* that enslaved individuals were persons, even as they denied them fundamental rights (Schechtman, 2014, pp. 125–126). Similarly, she argues that debates over fetal personhood often conflate moral status with metaphysical personhood. In her view, societies permitting abortion do not necessarily deny fetal personhood; rather, they accept that recognizing personhood is compatible with taking actions such as abortion (Schechtman, 2014, p. 128).

But does this defense succeed? Despite Schechtman's careful argumentation, we contend that her efforts to establish the PLV as non-conventionalist ultimately fall short. First, her over-inclusivity examples may not fully support her case. While she dismisses the idea of a talking poodle, higher primates have demonstrated significant communicative abilities. If, hypothetically, a group of primates could achieve a 90% success rate in language-based interaction, would it truly be irrational to feel disappointment if 'our' primate fell within the remaining 10%? Moreover, developmental expectations—such as the ability to read or write—are often shaped by social convention rather than essential criteria of personhood. This raises doubts about whether Schechtman's *default expectation* is a reliable foundation. Second, her under-inclusivity defense does not fully address the deeper issue of conventionalism. While she distinguishes metaphysical personhood from moral status, this does not rule out conventionalism. To do so, she would need to

demonstrate that it is rationally inconceivable to view Carla—or a fetus—as a non-person. This, we argue, remains unproven.

Ultimately, Schechtman's reliance on an evolutionary and social-institutional basis for person-space suggests that personhood is deeply ingrained but not *determined* by nature alone. In challenging cases like the dementia dilemma, our ingrained responses do not resolve the problem but rather reveal that personhood remains, at least in part, a matter of choice. Despite her efforts, the PLV still carries an element of personhood conventionalism—that is, personhood remains tethered to historically and culturally contingent practices rather than being an objective metaphysical fact.

## Concluding remarks

The question of whether to uphold advance directives in cases of severe cognitive decline touches on fundamental issues of autonomy, identity, and convention. Our analysis suggests that the authority of *S* at time *t* over their future self, *y* at time *t**, cannot be justified solely by an appeal to personal identity over time without invoking some degree of conventionalism. This, in turn, offers a strong argument for a conventionalist understanding of personal identity. From this perspective, advance directives derive their authority not because *S* at *t* remains unequivocally the same person as *y* at *t**, but because society has collectively chosen to treat them as integral to our conception of personhood and identity. This socially established framework allows us to view *y* at *t** as a continuation of *S* at *t*, making personal identity over time a product of shared practices—including advance directives. This point becomes particularly clear in cases involving durable power of attorney for healthcare, where the appointed agent is tasked not with making decisions based on *y*'s immediate needs, but rather with honoring the prior wishes of *S*.

These arguments, however, also open the door to an alternative interpretation: that advance directives may ultimately serve the interests of those who create them more than those who later exist in a diminished cognitive state. Some of the theoretical approaches we reviewed risk obscuring the fact that, in dementia-related cases, advance directives may provide greater reassurance to those drafting them than to the future individuals they are meant to govern. Furthermore, the extent to which their value is shaped by cultural biases and the stigma surrounding cognitive decline remains an open question. The key takeaway from our analysis, then, is that society, healthcare providers, and those drafting advance directives must critically examine the underlying assumptions about personal identity on which these practices depend. Acknowledging their conventionalist foundations does not diminish their importance but rather clarifies their purpose and limitations, allowing for more informed and deliberate decision-making by everyone involved.

## References

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics* (8th ed.). Oxford University Press.

Benzenhöfer, U., & Hack-Molitor, G. (2009). Luis Kutner and the development of the advance directive (living will). *GWAB*. https://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/34515

Braddon-Mitchell, D., & Miller, K. (2004). How to be a conventional person. *The Monist, 87*(4), 457–474.

Buchanan, A. E., & Brock, D. W. (1990). *Deciding for others: The ethics of surrogate decision-making*. Cambridge University Press.

DeGrazia, D. (1999). Advance directives, dementia, and 'the someone else problem'. *Bioethics, 13*(5), 373–391.

DeGrazia, D. (2005). *Human identity and bioethics*. Cambridge University Press.

Dresser, R. (1986). Life, death, and incompetent patients: Conceptual infirmities and hidden values in the law. *Arizona Law Review, 28*(3), 373–405.

Dresser, R., & Robertson, J. (1989). Quality of life and non-treatment decisions for incompetent patients: A critique of the orthodox approach. *Law, Medicine & Health Care, 17*(3), 234–244.

Jaworska, A. (2017). Advance directives and substitute decision-making. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2017 Edition). https://plato.stanford.edu/archives/sum2017/entries/advance-directives/

Kuczewski, M. (1994). Whose will is it anyway? A discussion of advance directives, personal identity, and consensus in medical ethics. *Bioethics, 8*(1), 27–48.

Luzio, H. (2025). Person conativist animalism. In A. Muñoz-Corcuera & N.-F. Wagner (Eds.), *Conventionalism about personal identity* (pp. 56–72). Routledge.

Macedo, J. C., et al. (2023). Perceptions, attitudes, and knowledge toward advance directives: A scoping review. *Healthcare, 11*(20), 2755.

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

Merricks, T. (2001). Realism about personal identity over time. In J. Tomberlin (Ed.), *Philosophical perspectives* (vol. 15: Metaphysics, pp. 173–187) [Supplement to *Noûs*, *35*(S15)]. Blackwell.

Parfit, D. (1984). *Reasons and persons*. Oxford University Press.

Schechtman, M. (2010). Personhood and the practical. *Theoretical Medicine and Bioethics, 31*(4), 271–283.

Schechtman, M. (2014). *Staying alive: Personal identity, practical concerns, and the unity of a life*. Oxford University Press.

Shoemaker, D. (2016). The stony metaphysical heart of animalism. In S. Blatti & P. F. Snowdon (Eds.), *Animalism: New essays on persons, animals, and identity* (pp. 303–328). Oxford University Press.

Wagner, N.-F. (2022). Personal identity, possible worlds, and medical ethics. *Medicine, Health Care and Philosophy, 25*(3), 429–437.