The True Self and Decision-Making Capacity

Article *in* American Journal of Bioethics · August 2024

DOI: 10.1080/15265161.2024.2361894

4 authors:

James Toomey
University of Iowa
14 PUBLICATIONS **12** CITATIONS
SEE PROFILE

Jonathan Lewis
The University of Manchester
51 PUBLICATIONS **363** CITATIONS
SEE PROFILE

Ivar R. Hannikainen
University of Granada
146 PUBLICATIONS **1,232** CITATIONS
SEE PROFILE

Brian D. Earp
University of Oxford
325 PUBLICATIONS **6,043** CITATIONS
SEE PROFILE

The American Journal of BIOETHICS

The True Self and Decision-Making Capacity

James Toomey, Jonathan Lewis, Ivar R. Hannikainen and Brian D. Earp[1]

This is a pre-print of an article currently in press. Please cite as:

Toomey, J., Lewis, J., Hannikainen, I. R., & Earp, B. D. (2024). The true self and decision-making capacity. *The American Journal of Bioethics,* in press.

[1] Corresponding author: brian.earp@gmail.com

Jennifer Hawkins (2024) offers two cases that challenge traditional accounts of decision-making capacity, according to which respect for a medical decision turns on an individual's cognitive capacities at the time the decision is made (Hawkins 2024; Appelbaum and Grisso 1988). In each of her described cases (involving anorexia nervosa and grief, respectively), a patient makes a decision that—although instrumentally rational at the time—does not reflect the patient's longer-term values due to being in a particular psychological state. Importantly, this state does not impair the patient's cognition, but rather predisposes them to make a decision that conflicts with their own broader values, beliefs, or desires.

Under traditional understandings of decision-making capacity, the patient's decision in either case must be followed by healthcare providers, insofar as it was made while in possession of the requisite cognitive abilities. But this, Hawkins suggests, is the wrong outcome. Although core cognitive capacities are *necessary* for decision-making capacity, they are not on her view *sufficient*. From her perspective, patients who clear the threshold of cognitive capacity are not entitled to have their decisions followed when there is good evidence they are making a serious prudential mistake while known to have a condition that makes people more likely than typical to make such mistakes.

As we interpret Hawkins's view, whether someone is making a "prudential mistake" depends in large part on aspects of who they are that are relatively stable and central, rather than fleeting or superficial. One way to interpret her cases, then, is that they involve patients who, although of typical cognitive functioning, are making decisions that do not ultimately reflect what some would be tempted to call their *true* self. In the psychological literature, the term "true self" refers to a lay conception of "who [someone] really is, deep down," as distinct from mere bodily continuity or numerical identity (Strohminger, Knobe, and Newman 2017). This concept, which psychologists and experimental philosophers have found explanatory in a number of contexts, corresponds to an everyday recognition that certain things people do and say are in some sense more true or authentic to who they are than are other things—captured, for instance, in the cliché that "it wasn't *really* him who said that horrible thing, it was the [anger, shame, illness, or what have you]" (see, e.g., Johnson, Robinson, and Mitchell 2004; Schlegel et al. 2013; Newman, Knobe, and Bloom 2014).

From this perspective, we might think that a patient with anorexia nervosa, whose "chances of recovery with treatment are still high," is not making a decision that reflects who he really is, deep down, when refusing to eat and "throwing away many years of life that would be quite good for him *from his perspective if he lived them*" (Hawkins 2024). Indeed, it might be thought that such a person isn't his true self when making such a decision. The same goes for someone who, in transient grief, declines life-sustaining treatment. Through the lens of the true self literature, then, Hawkins's proposal might be read as suggesting the following: that decisions made by a patient who meets the cognitive threshold of capacity need not necessarily be respected when these decisions—due to certain aberrant psychological states—do not accord with the patient's true self.[1]

We recently conducted a set of studies into factors that influence ordinary people's judgments about medical decision-making, including whether a patient's expressed treatment preference ought to be followed (Earp, Latham, and Tobia 2020; Earp et al. 2023; Toomey et al. 2024). In these studies, beliefs about the "true self" played a role. And they did so in a way that

[1] To be clear, we offer this as a friendly, approximate lens through which one might interpret Hawkins's proposal—naturally, her proposed test involves laudable analysis of additional procedural limitations that are not our focus here. And it might be that the concept of "prudential mistake" is distinct from "mistake from the perspective of the true self." But we would suggest, at least, that the true self might help explain Hawkins's intuitions about the cases she describes.

both complements and complicates Hawkins's principled argument. On the one hand, with Hawkins, the evidence we have so far collected suggests that many ordinary people take the continuity of the true self (or whether a treatment decision is seen to reflect a patient's true self), to be relevant when determining whether or not an expressed preference ought to be followed. Moreover, in Toomey et al. (2024), we found that US participants' views on whether a patient was their true self when they expressed a medical treatment preference correlated with endorsement of following that decision.

On the other hand, evidence from our studies suggests that ordinary people may not share Hawkins's view that cognitive capacity is *necessary* for a medical decision to be entitled to respect (Earp et al. 2023; Toomey et al. 2024). Rather, we have found a widespread intuition consistent with what might be thought of as the *inverse* of Hawkins's principle. Whereas Hawkins would not follow decisions above the threshold of cognitive capacity that do not reflect the patient's true self, ordinary people in our studies report endorsing decisions made *below* the threshold of cognitive capacity, insofar as those decisions reflect the patient's true self. In other words, perceived consistency with a patient's true self, rather than a certain level of cognitive capacity, seemed to be both necessary and sufficient for the participants in our studies to conclude that a decision ought to be respected.

In Toomey et al. (2024), we presented participants with a vignette involving a character with substantial cognitive decline, who would lack capacity under most traditional accounts, and who expressed a medical treatment preference—either for treatment or non-treatment. Most participants followed the character's expressed preference, in line with their perception that the character remained their true self, notwithstanding the lack of cognitive capacity. Again, this finding suggests that many ordinary people do *not* find patients' cognitive capacity to be necessary for their medical treatment preferences to be entitled to respect. Rather, our participants seemed perfectly comfortable endorsing decisions made with limited cognitive abilities (i.e., stating that they should be followed), as long as those decisions seemed to them to be consistent with the patient's true self. Thus, participants clearly believed that an individual's true self can, but does not necessarily, continue through their cognitive decline.

Of course, from a normative perspective, Hawkins may be right that considerations of a person's longer-term identity and deeper values should only come into play in deciding whether to follow their decision-making when it takes place above, but not below, the conventional

threshold of cognitive capacity. The common lay view to the contrary captured in our studies may well be wrong. But there *is* something parsimonious about a theory that would juxtapose Hawkins's intuitions in her cases with our findings regarding US participants. Specifically, it might be the authenticity of the decision to the true self (or whether the decision is a mistake from the perspective of who the person really is deep down) that matters in whether a decision ought to be respected in *all* cases, both above and below the traditional cognitive threshold.

If an account along these lines is right, the normative force of an expressed decision would simply always be a function of its relationship to the decision-maker's true self. Under Hawkins's account, a decision's normative force is instead sometimes a function of its relationship to a decision-maker's true self (as we are interpreting it), and sometimes a function of other cognitive factors. At a mini- mum, then, her view requires a theory of why certain factors are relevant in some cases but not others. More broadly, Hawkins might consider whether looking to the true self in all cases could explain her intuition that facts about cognition sometimes do seem to matter in capacity cases— after all, it might be that the persistence of the true self itself turns on facts about cognition, if not necessarily the same facts as traditional accounts. For example, it might turn on whatever the cognitive prerequisites are for maintaining a stable, normatively thick personal identity (i.e., true self) through time.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Appelbaum, P. S., and T. Grisso. 1988. Assessing patients' capac- ities to consent to treatment. *New England Journal of Medicine* 319 (25):1635–8. doi:10.1056/NEJM198812223192504.

Earp, B. D., I. R. Hanikainen, S. Dale, and S. R. Latham. 2023. Experimental philosophical bioethics, advance di- rectives, and the true self in dementia. In *Advances in experimental philosophy of medicine*, ed. A. De Block and K. Hens, 259–284. London: Bloomsbury.

Earp, B. D., S. R. Latham, and K. P. Tobia. 2020. Personal transformation and advance directives: An experimental bioethics approach. *The American Journal of Bioethics* 20 (8):72–5.

Hawkins, J. 2024. Affect, values and problems assessing decision-making capacity. *The American Journal of Bioethics.*

Johnson, J. T., M. D., Robinson, and E. B. Mitchell. 2024. Inferences about the authentic self: When do actions say more than mental states. Journal of Personality and Social Psychology 87 (5):615-30. doi:10.1037/0022-3514.87.5.615

Newman, G., J. Knobe, and P. Bloom. 2014. Value judgments and the true self. *Personality & Social Psychology Bulletin* 40 (2):203–16.

Schlegel, R. J., J. A. Hicks, W. E. Davis, K. A. Hirsch, and C. M. Smith. 2013. The dynamic interplay between perceived true self-knowledge and decision satisfaction. *Journal of Personality and Social Psychology* 104 (3):542– 58. doi:10.1037/a0031183.

Strohminger, N., J. Knobe, and G. Newman. 2017. The true self: A psychological concept distinct from the self. *Perspectives on Psychological Science* 12 (4):551–60. doi: 10.1177/1745691616689495.

Toomey, J., J. Lewis, I. R. Hannikainen, and B. D. Earp. 2024. Advance medical decision-making differs across first- and third-person perspectives. *The American Journal of Bioethics Empirical Bioethics*:1–9. doi:10.1080/23294515.2024.2336900.