

# Personal transformation and advance directives: an experimental bioethics approach

Brian D. Earp, Stephen R. Latham, & Kevin P. Tobia  
Yale University

This is a preprint of an accepted paper, now in press.

Earp, B. D., Latham, S. R., & Tobia, K. P. (2020). Personal transformation and advance directives: an experimental bioethics approach. *American Journal of Bioethics*, in press.

## Précis

Some people with dementia are transformed by the disease, to the point that family members may describe them as a “different person.” These transformations may be negative or positive. What factors affect the judgements of ordinary people about whether an advance directive (AD) should be followed in such cases? We conducted three studies to test the influence of (1) positive versus negative transformation and (2) “treat” versus “withhold treatment” AD on the judgements of US participants ( $n = 1676$ ) as to whether the AD should be followed and the extent to which the late-stage dementia patient had become a different person. We found that participants generally endorsed following the AD, irrespective of condition, but much less so when the patient had a positive transformation. Participants also favored “treat” over “withhold treatment” ADs. Unexpectedly, we found that AD type affected “different person” judgments: participants registered significantly weaker agreement with the proposition that the patient had become a different person when the AD instructed treatment. We discuss these results in the context of Walsh’s (2020) newly proposed normative model for AD decision-making.

**Keywords:** advance directives, personal identity, transformative experience, dementia

Walsh (2020) notes that dementia can inspire both personally and cognitively transformative experiences (Paul 2014), resulting in major changes to one's beliefs, values, preferences, and overall demeanor. Sometimes these changes are so pronounced that family members describe a late-stage dementia patient as a "different person" from the one they knew before the onset of the disease (e.g. Strohminger & Nichols 2015, p. 1470). Walsh argues that in these or similar "transformative" cases, clinicians should not assign decisive weight to advance directives. Instead, they should look to the current preferences or best interests of the patient.

Walsh's account is robust across different types of transformations. For example, Walsh acknowledges that dementia-related changes can be both positive and negative: some people with dementia seem to transform into happier, more sociable, or more child-like persons, while others seem to become meaner, more irritable, or more withdrawn. But the central claim applies to both types of cases: insofar as a transformative experience has occurred, it is not relevant (in itself) whether the change was for the better or worse.

The account also seems invariant across different types of advance directives (ADs). Walsh discusses ADs expressing a preference for *treatment* (e.g., Mr. O'Connor) and others expressing a preference for *withholding of treatment* (e.g., Mrs. Black). On Walsh's account, it seems that this distinction, too, should not affect the weight assigned to the AD versus the revealed preferences of the dementia patient post-transformation.

We were interested to learn whether ordinary people share this perspective, or whether their judgments about whether an AD should be followed might indeed be influenced by either of these factors (the positive versus negative direction of personal change, or the treatment-related content of the AD).

Why care about the judgments of ordinary people in this context? One reason to care is that they are the ones who must make real-life decisions about whether to follow an AD. If it turns out that their judgments are sensitive to factors that should be irrelevant on Walsh's (2020) account, we might try to find ways to address such sensitivity. Alternatively, it might be that Walsh's account should be updated to better accommodate the normative intuitions of ordinary people. Either way, it will be important to get a handle on the factors that influence such intuitions in the first place.<sup>1</sup>

---

<sup>1</sup> This approach may also help us understand why certain examples used to support a philosophical account such as Walsh's feel "intuitive" (or otherwise).

Previous work in psychology suggests that the first proposed factor (positive vs. negative change) might matter. Empirical studies show that when someone undergoes a sufficiently salient personal transformation, especially one that affects their moral attributes, they are likely to be regarded as “not the same person” as the one who existed prior to the transformation (Strohming & Nichols 2014, 2015; see generally Shoemaker & Tobia, forthcoming). However, there is an asymmetry: moral deteriorations are typically seen as more identity-disruptive than moral improvements of a similar magnitude (Tobia 2015, 2016; Earp et al. 2019). Accordingly, we might expect that the positive versus negative direction of change brought about by dementia could affect the extent to which a late-stage dementia patient is seen as a “different person” to the one who signed the AD.<sup>2</sup> This, in turn, may affect how much weight people think the AD should get.

To explore these questions, we conducted three experimental studies. Our aim was to test the influence of (1) a positive versus negative personal transformation (brought about by dementia), and (2) a “treat” versus “withhold” AD on ordinary people’s judgments about whether the AD should be followed—and on whether the late-stage dementia patient is now a “different person.” The complete methods, materials, results, participant demographics, and open data are on the Open Science Framework: <https://osf.io/m6tgi/>.

In our first study, we examined the judgments of US American participants ( $n = 344$ )<sup>3</sup> regarding “Robin,” who signs an AD requesting either treatment or withholding of treatment. In the story, Robin then either morally improves or deteriorates as a result of dementia, and then contracts pneumonia, becoming lethargic and unresponsive. The doctor makes clear that the illness is curable, but if left untreated will cause Robin’s death. We asked participants, in each condition, first whether the doctor *should follow the AD* (assuming the family agrees with any decision), and then whether they agreed that the late-stage dementia patient was *no longer the same person* as Robin.

The results are depicted in Figure 1. We found that while most participants endorsed following the AD regardless of condition, such endorsement was stronger when the AD

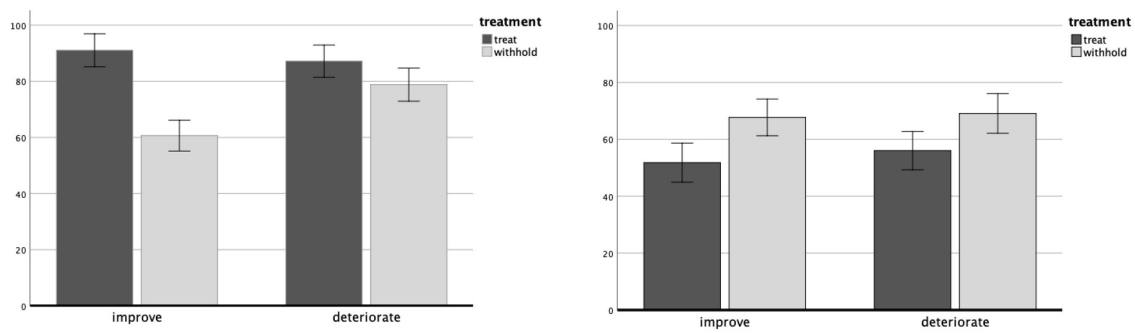
---

<sup>2</sup> There are important theoretical questions about the meaning and philosophical significance of ordinary people’s judgments about the “different person” relation—for example, do they concern a break in *numerical identity* or *qualitative similarity* (see, e.g., Dranseika 2017; Starmans & Bloom 2018)? Plausibly, same/different person judgments are context-sensitive and depend, among other things, on the purpose(s) for which they are being elicited.

<sup>3</sup> Participants for all studies were recruited through Amazon’s Mechanical Turk (MTurk) platform. Sample sizes reported in this paper are the final  $n$  after excluding participants who failed one or more of three embedded attention checks.

instructed treatment than withholding of treatment. Participants also more strongly endorsed following the AD when Robin deteriorated than when she improved. Finally, there was a significant interaction between these factors. When Robin improved, average ratings for following a “withhold” AD were *much* lower than average ratings for following a “treat” AD. However, when Robin deteriorated, these ratings were less divergent.

With respect to identity judgments, there was a significant effect of AD: participants more strongly agreed that Robin was a “different person” when the AD instructed withholding of treatment than when it instructed providing treatment. Contrary to our expectations, there was no effect of improvement versus deterioration and no interaction (see supplemental materials for statistical analyses).



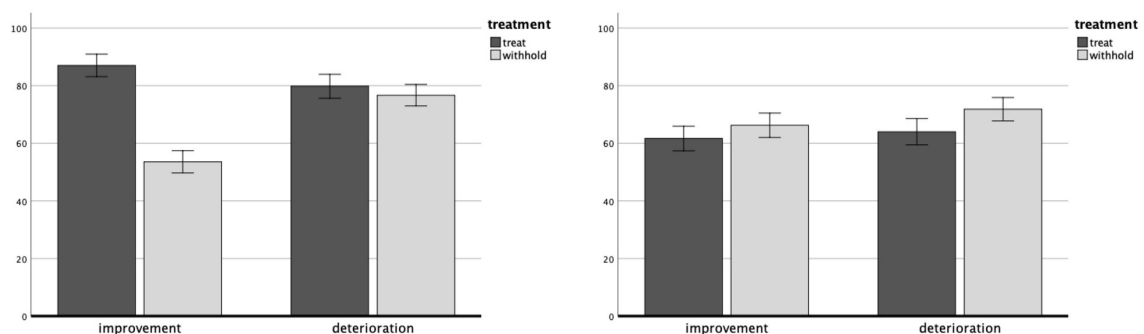
**Figure 1.** Study 1 results for “follow the AD” judgments (left panel), showing mean agreement ratings concerning whether the doctor should follow the AD, and for “different person” judgments (right panel), showing mean agreement ratings concerning whether the late-stage dementia patient is a different person from Robin. 100 = strongest agreement with each proposition. Error bars indicate 95% confidence intervals.

Altogether, the results from Study 1 indicate that the judgments of ordinary people about the weight to assign to a pre-transformation AD are significantly affected both by the content of the AD (whether it instructs treatment or withholding of treatment) *and* by the positive or negative direction of personal change brought about by dementia.

Walsh’s (2020) normative model seems to suggest that neither of these factors should matter in and of itself. Rather, the AD should be discounted – whatever its content – insofar as the preferences or best interests of the post-transformation patient have changed. However, it might be that the factors matter indirectly. In our study, we did not explicitly state that “Robin” had updated preferences that were different from (or consonant with) the instructions of the AD. So perhaps participants were influenced by a “best interests” intuition, like the medical staff in the case study of Mrs. Black (the woman

who signed a “withhold” AD but underwent a positive transformation). As Walsh describes, the staff were “devastated” by having to withhold treatment from this “exceptionally happy” older dementia patient. Likewise, in our study, participants were most reluctant to follow the AD when it instructed withholding treatment after “Robin” had changed for the better.

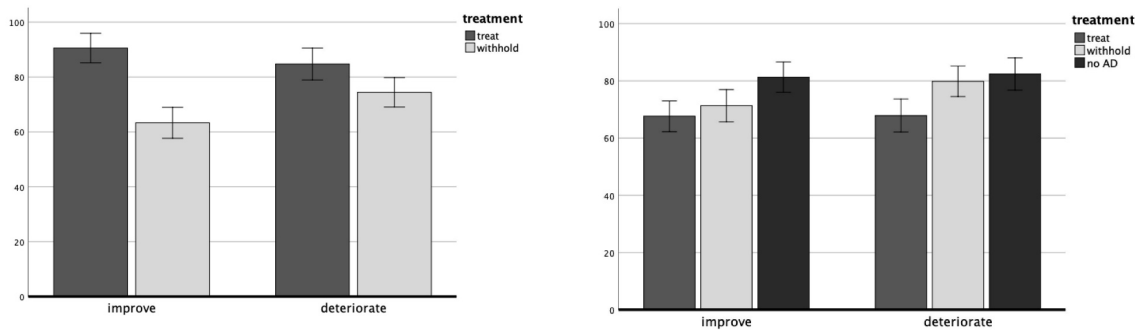
To confirm these results, we conducted a second study ( $n = 755$ ) based more closely on the story of Mrs. Black as described by Walsh (2020) (see supplemental materials). We replicated the pattern of results: participants were least inclined to follow the “withhold” AD—especially when the patient had a positive transformation (see Figure 2, left panel).



**Figure 2.** Study 2 results for “follow the AD” judgments (left panel), and for “different person” judgments. Error bars represent 95% confidence intervals.

We also replicated the “different person” finding from Study 1. Participants more strongly agreed that the patient and the signer of the AD were not the same person when the AD instructed withholding of treatment.<sup>4</sup> Thus, it appears that making a “different person” judgment *in the context of AD decision-making* affects participant responses (i.e., depending on what the AD instructs). To explore this, we conducted a final study ( $n = 577$ ). The design of our final study was the same as the first, except that we added a control condition for both the “improve” and “deteriorate” scenarios in which no mention was made of an AD or AD-related decision-making, and *only* the question about being a “different person” was asked. Results are depicted in Figure 3.

<sup>4</sup> How might this finding be explained? We know that participants prefer “treat” over “withhold” ADs. Perhaps in their reluctance to see treatment withheld from a patient who could easily be cured—especially if the patient has undergone a positive transformation—they find themselves inclined to judge that the patient and the person who signed the “withhold” AD are, after all, different people.



**Figure 3.** Study 3 results for “follow the AD” judgments (left panel), and “different person” judgments (right panel). Error bars indicate 95% CIs.

First, in the conditions dealing with ADs, we found a similar pattern of results for following the AD as we observed in the previous studies (see left panel). Such results are therefore robust. And with respect to the “different person” question (right panel), we were able to confirm that asking this question in the context of an AD decision did make a difference to the findings. Specifically, agreement that the post-transformation patient was a different person was highest when there was no AD mentioned at all, next highest when there was a “withhold” AD, and lowest when there was a “treat” AD (see supplemental materials for full results).

These exploratory studies, inspired by Walsh’s important article, highlight factors that appear to influence ordinary people’s judgments about advance directives in cases of personal transformation. First, the *content* of the AD matters. Holding everything else equal, participants more strongly judge that a clinician should follow an AD when it instructs treatment compared to withholding of treatment. We also find that *direction of change* matters: When a transformation is positive, participants strongly prefer a “treat” over a “withhold” AD, but when it is negative, this difference is much smaller. These results provide initial evidence that the view of ordinary people diverges from Walsh’s account, according to which neither of these factors should affect the weight that is assigned to a pre-transformation AD (per se).

We also hope to have illustrated the usefulness of an experimental bioethics or “bioxphi” approach (Earp *et al.* 2020). This approach can help identify factors that influence ordinary people’s ethical judgments—and reveal where these judgments diverge from what normative models recommend.

## References

- Dranseika, V. (2017). On the ambiguity of 'the same person.' *AJOB Neuroscience*, 8(3), 184–186. <https://doi.org/10.1080/21507740.2017.1366588>
- Earp, B. D. *et al.* (2020). Experimental philosophical bioethics. *AJOB: Empirical Bioethics* 11(1), 30-33. <https://doi.org/10.1080/23294515.2020.1714792>
- Earp, B. D., Skorburg, J. A., Everett, J. A. C., & Savulescu, J. (2019). Addiction, identity, morality. *AJOB: Empirical Bioethics*, 10(2), 136–153. <https://doi.org/10.1080/23294515.2019.1590480>
- Paul, L. A. (2014). *Transformative experience*. Oxford: Oxford University Press.
- Shoemaker, D. & Tobia, K. (forthcoming). Personal identity. *Oxford Handbook of Moral Psychology*.
- Starmans, C., & Bloom, P. (2018). Nothing personal: What psychologists get wrong about identity. *Trends in Cognitive Sciences*, 22(7), 566-568.
- Strohming, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131(1), 159–171. <https://doi.org/10.1016/j.cognition.2013.12.005>
- Strohming, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, 26(9), 1469–1479. <https://doi.org/10.1177/0956797615592381>
- Tobia, K. (2015). Personal identity and the Phineas Gage effect. *Analysis*, 75(3), 396–405. <https://doi.org/10.1093/analys/anv041>
- Tobia, K. (2016). Personal identity, direction of change, and neuroethics. *Neuroethics*, 9(1), 37–43. <https://doi.org/10.1007/s12152-016-9248-9>
- Walsh, E. (2020). Cognitive transformation, dementia, and the moral weight of advance directives. *American Journal of Bioethics*, in press.