

Letter

If You Become Evil,
Do You Die?Christina Starmans^{1,*} and
Paul Bloom²

De Freitas *et al.* [1] agree with us about the importance of distinguishing between personal identity and similarity. We agree with them that individuals can be obliterated through severe neurodegeneration and the like: as we put it in our original article [2], ‘There are cases . . . where it may be thought that a person ceases to exist while their body survives, as in severe dementia.’ [3]. Finally, we share an appreciation of research on the notion of a morally good true self (indeed, one of us is a coauthor on the first paper on the topic).

However, our disagreement is a major one. We think that when someone becomes immoral, people see them as undergoing a substantial transformation. But the person does not cease to exist; we think the answer to the question ‘If you become evil, do you die?’, is plainly ‘no’. The arguments made by De Freitas *et al.* suggest that they think that the answer is ‘yes’. They argue that moral goodness is seen as intrinsic to a person’s existence: ‘Eliminate that quality, and you eliminate the person’. If Bob loses his goodness, they say, then ‘people believe he is no longer there’.

For them, this follows from the true self findings they review [4–6]. This research explores what people see as the most important, essential, or central features of an individual. But we see this as being distinct from the project of determining the features that lead individuals to persist over time. There is a world of difference between thinking that the most important, essential, or central feature of Bob is his kindness, and thinking that if Bob were to lose his kindness, he would cease to exist.

De Freitas *et al.* point out that true self research reveals an asymmetry – it is seen as more of a change when someone goes from good to bad than when someone goes from bad to good [4–7]. We agree, and indeed we are happy to endorse their own account of this, which is that when someone becomes good, people believe that this reflects the influence of an already existing true self and is thus less of a transformation. However, none of this is a challenge to our view that these transformations preserve personal identity.

On a minor point, De Freitas *et al.* also insist that we take literally the language people use to describe dramatic moral changes. They argue that when Gage was described as ‘no longer Gage’, ‘presumably [Gage’s family] were not just casually indicating that he was no longer a nice guy’. We agree that there is nothing casual here about this type of statement – brain damage is serious business – and presumably the family meant something broader, akin to ‘Gage no longer has the important traits that we’ve always associated with him’. But we see this as similar to saying, ‘I’m just not myself today’, which obviously cannot be meant literally, and illustrates that we typically use this type of language to talk about changes, not obliteration.

Finally, we agree that it is important to distinguish between personal and legal identity; there are cases where a legal notion (ownership, consent, culpability) is different from the psychological one. However, legal identity is at least partially based on intuitions about personal identity. As a science fiction example, imagine that Bob dies and his body is donated to science. Fred, sound of mind but poor of body, has his brain transplanted into Bob’s body. In this case we would certainly assign the person who looks like Bob – the body that used to be Bob’s – a new name and a new legal identity. Why

then do we not do this when a person becomes immoral? We believe, in this type of case, that nobody thinks there is a numerically different person. Instead, the perception is that there is one person who has changed dramatically.

We conclude with some street corner experimental philosophy, asking our readers this: have you ever encountered someone, either in real life or in fiction, who started off good and then become immoral? If so, did the person then disappear? Did their body become a shell, now occupied by a different individual? We take it that the answer is ‘no’. De Freitas *et al.* might say that we are being unfair – they are not saying that individuals who become immoral literally cease to exist. But we do not know how else to interpret their strong claims, such as ‘eliminate the person’ and ‘he is no longer there’. If they do mean these claims metaphorically, as ways of talking about major personal changes, then we are not sure why they see themselves as disagreeing with us.

¹Department of Psychology, University of Toronto, Toronto, ON, Canada

²Department of Psychology, Yale University, New Haven, CT, USA

*Correspondence:

christina.starmans@utoronto.ca (C. Starmans).
<https://doi.org/10.1016/j.tics.2018.05.009>

References

1. De Freitas, J. *et al.* (2018) Moral goodness is the essence of personal identity. *Trends Cogn. Sci.*
2. Starmans, C. and Bloom, P. (2018) Nothing personal: what psychologists get wrong about identity. *Trends Cogn. Sci.* Published online May 1, 2018. <http://dx.doi.org/10.1016/j.tics.2018.04.002>
3. Strohminger, N. and Nichols, S. (2015) Neurodegeneration and identity. *Psychol. Sci.* 26, 1469–1479
4. Newman, G.E. *et al.* (2015) Beliefs about the true self explain asymmetries based on moral judgment. *Cogn. Sci.* 39, 96–125
5. De Freitas, J. *et al.* (2017) Normative judgments and individual essence. *Cogn. Sci.* 41, 382–402
6. De Freitas, J. *et al.* (2017) Origins of the belief in good true selves. *Trends Cogn. Sci.* 21, 634–636
7. Tobia, K.P. (2015) Personal identity and the Phineas Gage effect. *Analysis* 75, 396–405