

## **Moral goodness is the essence of personal identity**

Julian De Freitas<sup>1</sup>, Mina Cikara<sup>1</sup>, Igor Grossmann<sup>2</sup>, & Rebecca Schlegel<sup>3</sup>

1 – Department of Psychology, Harvard University

2 – Department of Psychology, University of Waterloo

3 – Department of Psychology, University of Texas A&M

[Words: 813]

Keywords: Self; True Self; Psychological Essentialism; Moral Cognition

Starmans and Bloom ([1]; henceforth S&B) argue that research on the centrality of morality in people's intuitions about personal identity does not reveal much about people's notions of personal identity (whether an individual is the same person at time<sub>a</sub> and time<sub>a+1</sub>), but only something about their notions of similarity (how much the person at time<sub>a</sub> shares properties with the person at time<sub>a+1</sub>). We agree with S&B that it is important to distinguish between these constructs but disagree with their conclusion. Here we briefly review evidence that judgments regarding personal identity following a change in moral character cannot be explained by a mere (dis)similarity account.

First, consider their thought experiment: "Suppose that when Bob was 20, he was the nicest of people. Generous, kind to animals— a real mensch. But then Bob experienced a profound moral transformation, and he turned into a terrible person: mean, selfish, psychopathic, a man who robs stores and kicks dogs". Now, invert their thought experiment: "Suppose that when Bob was 20, he was a terrible person: mean, selfish, psychopathic, a man who robs stores and kicks dogs. But then Bob experienced a profound moral transformation, and he turned into the nicest of people. Generous, kind to animals— a real mensch." Comparing this moral improvement scenario to the moral deterioration scenario offered by S&B, we see that the magnitude of change is the same (i.e., dissimilarity is equated). Nonetheless, we and others consistently find that in such improvement scenarios people say that Bob is still the *same* person (reviewed by [2]). Why? Because people believe that Bob's true self, his essence, consists of the morally *good* traits (e.g., [3–4]). Thus, when Bob undergoes significant moral improvement, people believe that his good essence has emerged, that he has 'found himself.' But when Bob undergoes significant moral deterioration, people believe that he is no longer there. Furthermore, we find via mediation analyses that participants' judgments of similarity between such cases do not explain identity judgments, whereas beliefs about a morally good essence do [5].

Second, we disagree that people in these experiments are merely speaking colloquially when they agree that the person is no longer present. When Phineas Gage's head was penetrated by a tampering iron, his family said he was "no longer Gage" [6]. Presumably they were not just casually indicating that he was no longer a nice guy, but really did feel that the *person* they knew and loved was no longer present. The same goes for patients who have undergone significant neurodegeneration [7]. We think that their relatives are saying more than just that their loved one has undergone a big psychological change. Rather, they are picking out something deeper: Despite there still being a superficial physical presence, this seems to be just a shell that is no longer inhabited by the *person* who they came to know and love.

Third, regarding passports, birthdays, and taxes, we think it is important to distinguish between personal identity and legal identity [8]. Even if people have the sense that someone is no longer the same person, the superficial characteristics of the person pertaining to their legal identity remain the same. At the very least, there continues to be a body with the same perceptual abilities, occupying the same spatiotemporally continuous path, with the same indisputable physical birthdate and history. So in one superficial sense, people are forced to admit, as a Capgras patient might even admit, that

the ‘person’ is still there. But in another sense — the sense that counts for personal identity — people cannot shake the feeling that the person they know is gone.

Finally, people’s beliefs about a good true self appear to be a form of psychological essentialism (PE; reviewed by [2]), which has previously been linked to category membership and identity using precisely the sorts of studies that S&B recommend [e.g., 9–10]. An important property of PE is that removing an entity’s seemingly essential characteristics is more disruptive to its identity than removing its seemingly peripheral characteristics. This is exactly the pattern observed in recent experiments: Removing morally good traits leads to a larger sense of disruption to personal identity compared to other kinds of traits, including morally bad traits of an equal magnitude. Furthermore, beliefs about a good true self show various hallmarks of psychological essentialism [11]. People believe that morally good traits are innate and cross-temporally stable, that there is a boundary separating the self-essence from other aspects of the self, and that self-essences have non-obvious properties and are diagnostic of what is true about an individual. Finally, like other documented effects of PE, the good true self belief seems to operate similarly across cultural and individual differences [12]. To our minds, the most parsimonious interpretation of these various findings is that people believe that moral goodness is the fundamental quality that defines the person. Eliminate this quality, and you eliminate the person.

## References

1. Starmans, C., & Bloom, P. (2018). Nothing personal: What psychologists get wrong about identity. *Trends in Cognitive Sciences*.
2. De Freitas, J., Cikara, M., Grossmann, I., & Schlegel, R. (2017). Origins of the belief in morally good true selves. *Trends in Cognitive Sciences*, 21(9), 534–636.
3. Newman, G. E., Bloom, P., & Knobe, J. (2014). Value judgments and the true self. *Personality and Social Psychology Bulletin*, 40(2), 203–216.
4. Newman, G. E., De Freitas, J., & Knobe, J. (2015). Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science*, 39(1), 96–125.
5. De Freitas, J., Tobia, K., Newman, J. E., & Knobe, J. (2017). Normative judgments and individual essence. *Cognitive Science*, 41, 382–402.
6. Harlow, J. 1868. Recovery from the passage of an iron bar through the head. *Publications of the Massachusetts Medical Society*, 2, 327–347.
7. Strohminger, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, 26(9), 1469–1479.
8. Tobia, K. P. (2015). Personal identity and the Phineas Gage effect. *Analysis*, 75(3),

396–405.

9. Gelman, S.A. (2003) *The essential child: Origins of essentialism in everyday thought*, Oxford University Press.
10. Gutheil, G., & Rosengren, K. S. (1996). A rose by any other name: Preschoolers' understanding of individual identity across name and appearance changes. *British Journal of Developmental Psychology*, *14*(4), 477–498.
11. Christy, A. G., Schlegel, R. J., & Cimpian, A. (2017). The essence of the individual: The pervasive belief in the true self is an instance of psychological essentialism. <http://doi.org/10.17605/OSF.IO/K3JBA>.
12. De Freitas, J., Sarkissian, H., Newman, G. E., Grossmann, I., De Brigard, F., Luco, A., & Knobe, J. (2017). Consistent belief in a good true self in misanthropes and three interdependent cultures. *Cognitive Science*.