

## Forum

Nothing Personal:  
What Psychologists  
Get Wrong about  
IdentityChristina Starmans<sup>1,\*</sup> and  
Paul Bloom<sup>2</sup>

**What makes someone the same person over time? There is a growing body of research exploring how people ordinarily think about personal identity. We argue here that many of the experiments in this domain fail to properly distinguish similarity from personal identity, and therefore certain conclusions regarding common-sense intuitions about identity are not supported.**

Bob and Zeke Kravitz are identical twin brothers in their 20s, who live together. Late one night, while Zeke sleeps, Bob leaves the house, gets drunk in a bar, and then robs a liquor store. Fleeing the scene, he falls onto the sidewalk and lands on his face, which is severely bruised by the time he gets home. When the police look at the surveillance footage, an officer recognizes the robber – “it’s one of those damn Kravitz boys!” – and they drive over to the twins’ home. When they arrive, Zeke looks exactly like the image on the store camera and Bob, with his banged-up face, does not.

Who should the police arrest? This is an easy question: even though Zeke might resemble the robber more, the police are really after Bob, since it was he who committed the crime. This example, inspired by real events [1], illustrates how intuitions about identity differ from intuitions about similarity. Under conditions of uncertainty, we use similarity as a cue to identity; this is what we do in everyday life when we see

someone and recognize them as our friend, our child, or our spouse. But similarity is a heuristic, and can go wrong, which is what would happen if the police arrested Zeke.

This example illustrates an important distinction between two types of identity. The first, which philosophers call numerical identity, is the sense in which a single thing persists over time, as when baby Bob is identical to adult Bob. If you once smiled at baby Bob, and later smiled at adult Bob, you have smiled twice at one person. The second is what philosophers refer to as qualitative identity: the sharing of properties. This is the sense in which two different things can be identical, as when baby Bob is identical to baby Zeke. Following standard practice, we will refer to numerical identity as ‘numerical identity’, ‘personal identity’, or simply ‘identity’, and to qualitative identity as ‘similarity’.

Intuitions about similarity are central to our psychological lives, and have been the focus of an immense body of research. But intuitions about personal identity are no less important, essential for all human interactions, at the foundation of promises and obligations, moral and legal responsibility, and emotions like gratitude, envy, and love. None of this would be possible without the capacity to understand that someone at Time 1 can be the same person at Time 2 even if they undergo physical and psychological changes. Increasingly, such intuitions are also objects of scientific study, as psychologists have become interested in uncovering how people ordinarily think about personal identity [2–7].

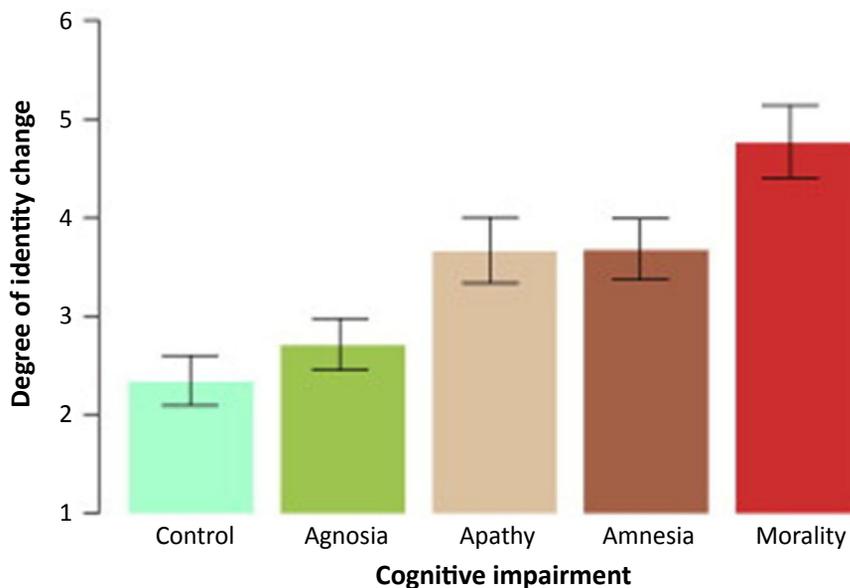
However, we believe there is a pervasive confusion in many of these experiments that causes the conclusions to be uninformative about how people think about personal identity. Consider one significant series of studies by Strohminger and Nichols [6] (henceforth S&N), exploring

which aspects of the mind are seen as central to personal identity. Their findings reveal an intriguing hierarchy (Figure 1). Lower-level perceptual traits, such as the ability to recognize objects, are least connected to identity; memory, especially autobiographical memory, is somewhat important, and most important of all are moral traits.

For S&N, these results testify to ‘the prevalence of moral traits in numerical identity perception’. We agree that this hierarchy captures something deep about people’s intuitions about the self. But we disagree with the conclusion drawn by S&N that these findings are about numerical identity. Instead, their methods tap intuitions about similarity. This concern applies to a number of papers investigating identity (e.g., [5,8]); we focus on S&N as an example because of the conceptual clarity, range of methods, and philosophical sophistication.

Consider their main conclusion, about the centrality of morality. This is plausible for similarity, but it is deeply unintuitive to see morality as pertaining to personal identity. 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. Would people believe that Bob ceased to exist, and that a new person came into existence?

We doubt it. After his moral change, nobody would search for Bob; they would not give away his property; his debts would not be forgiven. The current Bob would not need to apply for a new passport; he would not celebrate his first birthday a year after he became evil. Bob’s parents and old friends would not approach Bob and say, seriously: ‘You look just like someone I cared about. What’s your name?’ Rather, they would say, ‘Bob, you used to be such a nice guy!



Trends in Cognitive Sciences

Figure 1. The Impact of Changes to the Brain on Identity Change, from Strohminger and Nichols [6].

What happened?’ Bob does not cease to exist; nobody is created anew.

Why, then, do S&N get the results that they do? Our concern is that their questions (and the questions used by many other researchers) are ambiguous between the two senses of identity (see also [9]). For example, in one study, S&N asked participants to rate the extent to

which a person who had taken a mind-altering drug was ‘the same person as before’ versus ‘completely different now’ (the same response measure was used in [8]). In another, they asked participants to rate the extent to which a (partial) brain transplant recipient was ‘still Jim’. These questions can indeed tap numerical identity, but they can also be about similarity. When someone has

changed significantly, one might say he is ‘completely different now’ and after such a big shift in Bob’s moral behavior, it sounds perfectly natural to say ‘he isn’t Bob anymore’. (A similar response measure was used in [5], where participants were asked whether, after a change, ‘I will really be myself . . . not at all be myself’). But this figure of speech is a way of saying that there has been significant psychological change, not that one person has ceased to exist and another has been created.

Thus, at best, we cannot tell whether these data reflect people’s intuitions about similarity or about numerical identity. But for the reasons described above, we think that the most natural reading of these questions leads participants to answer in terms of similarity, not personal identity. In the real world, nobody sees moral change as influencing identity. This does not diminish the importance or interest of these data; it is fascinating that moral changes are seen as a much larger change to a person’s self than changes in memory or desires. But we wish to encourage future work to explore the question of personal identity more directly.

How can we experimentally distinguish intuitions about similarity from intuitions about identity? One method is to draw upon a much-studied topic in cognitive psychology and developmental psychology (see, e.g., [10,11]), and turn to the understanding of numerical identity in objects (see Box 1 for further discussion). Here the intuitions are clear. If George and Amy both show up to a talk carrying indistinguishable white coffee mugs with daffodils on them, we have a case of similarity but not numerical identity. However, if George lends his mug to Amy, and she paints roses over the daffodils and then returns it to George, we have a case of numerical identity but not similarity. These easier-to-understand examples

#### Box 1. Psychological Identity and Object Identity

We argue that changes to a person over time, even those that lead to significant dissimilarities, do not necessarily lead people to perceive a change in personal identity. But we are not denying a central premise of much contemporary work in this domain (including the studies we are critiquing), which is that there are multiple types of identity. In our initial example of Bob Kravitz, personal identity corresponded to Bob’s body: one tracks Bob from home to the liquor store and back again in the same way that one would track a cup or chair or computer. But it’s long been known that ‘same person’ and ‘same object’ can be psychologically dissociated. Examples include beliefs about reincarnation, demonic possession, and ‘body swaps’ as in John Locke’s parable of the prince and the cobbler, or the movie *Freaky Friday*.

Such intuitions might arise from intuitive dualism: the conception of a person as an immaterial soul that can hop from one body to another. But even those who reject dualism can dissociate personal identity from physical identity. There are cases (discussed in detail in [7]) where it may be thought that a person ceases to exist while their body survives, as in severe dementia. Conversely, there are hypothetical cases like the *Star Trek* transporter, in which someone’s body is obliterated and a perfect duplicate of that person is created out of new material. Despite the lack of spatiotemporal continuity, many people, including neuroscientists, philosophers, and the first author of this paper, believe identity is retained – though many others, such as the second author of this paper, do not.

can be used to distinguish the two separate meanings of identity for participants, who can then be asked about the types of changes explored in S&N and elsewhere.

We see this as a friendly correction; the finding by S&N and others about the centrality of morality to the self is fascinating, and connects well to other research on the 'moral true self' [8,12]. It's just not about personal identity.

<sup>1</sup>Department of Psychology, University of Toronto, Ontario, Canada

<sup>2</sup>Department of Psychology, Yale University, New Haven, CT, USA

**\*Correspondence:**

christina.starmans@utoronto.ca (C. Starmans).

<https://doi.org/10.1016/j.tics.2018.04.002>

**References**

1. Pinker, S. (1997) *How the Mind Works*, Norton
2. Bartels, D.M. and Urminsky, O. (2011) On intertemporal selfishness: how the perceived instability of identity underlies impatient consumption. *J. Consum. Res.* 38, 182–198
3. Blok, S. et al. (2005) Individuals and their concepts. In *Categorization Inside and Outside the Laboratory: Essays in Honor of Douglas L. Medin (8232)* (Ahn, W.K., ed.), In pp. 127–149, American Psychological Association
4. Chen, S.Y. et al. (2016) Beliefs about the causal structure of the self-concept determine which changes disrupt personal identity. *Psychol. Sci.* 27, 1398–1406
5. Molouki, S. and Bartels, D.M. (2017) Personal change and the continuity of the self. *Cogn. Psychol.* 93, 1–17
6. Strohminger, N. and Nichols, S. (2014) The essential moral self. *Cognition* 131, 159–171
7. Strohminger, N. and Nichols, S. (2015) Neurodegeneration and identity. *Psychol. Sci.* 26, 1469–1479
8. De Freitas, J. et al. (2017) Normative judgments and individual essence. *Cogn. Sci.* 41, 382–402
9. Dranseika, V. (2017) On the ambiguity of 'the same person'. *AJOB Neurosci.* 8, 184–186
10. Xu, F. and Carey, S. (1996) Infants' metaphysics: the case of numerical identity. *Cogn. Psychol.* 30, 111–153
11. Gelman, S.A. et al. (2016) Children seek historical traces of owned objects. *Child Dev.* 87, 239–255
12. De Freitas, J. et al. (2017) Origins of the belief in good true selves. *Trends Cogn. Sci.* 21, 634–636