THE FORUM

Gauging Personal Identity

Kevin Tobia on how our intuitions about personal identity reflect moral norms

Philosophers have long contemplated what conditions are required to maintain personal identity over time. Some have argued that identity over time might be broken by large and extreme changes, such as permanent and total memory loss or radical personality change. Consider a vignette based on the well-known story of Phineas Gage, a nineteenth-century railroad worker:

> Phineas is extremely kind; he really enjoys helping people. He is also employed as a railroad worker. One day at work, a railroad explosion causes a large iron spike to fly out and into his head, and he is immediately taken for emergency surgery. The doctors manage to remove the iron spike and their patient is fortunate to survive. However, in some ways this man after the accident is remarkably different from Phineas before the accident. Phineas before the accident was extremely kind and enjoyed helping people, but the man after the accident is now extremely cruel; he even enjoys harming people.

As the story famously goes, Phineas's friends and family saw the man after the accident as 'no longer Gage'. Though aspects of the historical account are contested, the story is often told as revealing that extreme changes seem to disrupt identity.

However, in the historical case, the accident involves not just a large change, but specifically a deterioration: the man after the accident is seen as worse than Phineas before the accident. The typical Phineas Gage interpretation is that the *magnitude* of a big change disrupts identity. But might this other feature, the *direction* of change ('improvement' or 'deterioration'), be partly responsible for judgements about identity?

To test this suspicion, I conducted a short experimental philosophy study. Participants saw either a 'deterioration scenario' (the example above), or the following 'improvement scenario', in which a comparably large change results in an improvement:

> Phineas is extremely cruel; he really enjoys harming people. He is also employed as a railroad worker. One day at work, a railroad explosion causes a large iron spike to fly out and into his head, and he is immediately taken for emergency surgery. The doctors manage to remove the iron spike and their patient is fortunate to survive. However, in some ways this man after the accident is remarkably different from Phineas before the accident. Phineas before the accident was extremely cruel and enjoyed harming people, but the man after the accident is now extremely kind; he even enjoys helping people.

Participants evaluated whether the (improved or deteriorated) man after the accident is still Phineas Gage. Compared to those participants who saw the classic 'deterioration' story, those who read about Phineas improving were more inclined to see him as still the same person. This reveals that big changes don't affect perceived identity equally; instead, how we see identity depends on the particular direction of change.

(For those who are interested, the exact 'numerical identity' question participants were asked is the following: 'Art and Bart disagree over what happened in this story. Art thinks that Phineas before the accident and the man after the accident are different in some respects but are still the same person. To Art, it seems like one person (Phineas) experienced some changes. Bart disagrees. He thinks that after the accident, the original man named Phineas does not exist anymore; the man after the accident is a different person. To Bart, it seems like one person died (Phineas before the accident), and it is really a different person entirely that exists after the accident (the man after the accident).')

The same effect arises when considering Parfit's famous Russian Nobleman case. Consider a vignette of Parfit's original thought experiment:

> In several years, a young Russian will inherit vast estates. Because he has socialist ideals, he intends, now, to give the land to the peasants. But he knows that in time his ideals may fade. To guard against this possibility, he does two things. He first signs a legal document, which will automatically give away the land, and which can be revoked only with his wife's consent. He then says to his wife, 'Promise me that, if I ever change my mind, and ask you to revoke this document, you will not consent'. He adds, 'I regard my ideals as essential to me. If I lose these ideals, I want you to think that I cease to exist. I want you to regard your husband then, not as me, the man who asks you for this promise, but only as his corrupted later self. Promise me that you would not do what he asks.'

Parfit suggests some might think of the older Russian as a different person from the younger Russian, noting that

> […] if this man's wife made this promise, and he did in middle age ask her to revoke the document, she might plausibly regard herself as not released from her commitment. It might seem to her as if she has obligations to two different people. She might believe that to do what her husband now asks would be a betrayal of the young man whom she loved and married. And she might regard what her husband now says as unable to acquit her of disloyalty to this young man.

Although Parfit does not himself claim the young and old Russian are numerically non-identical, his Russian Nobleman case is a seminal thought experiment sometimes cited as evidence for the view that major dissimilarities appear to break personal identity. However, this judgement might also gain its force from a Phineas Gage effect, as the change described may be seen as not only a big change, but specifically a deterioration.

To test this, I ran a second experiment. Participants read either the original Russian Nobleman Case as above, which is a 'deterioration case', or a slightly revised 'improvement' case:

> In several years, a young Russian will inherit vast estates. Because he has anti-socialist ideals, he intends, now, to not give the land to the peasants. But he knows that in time his

> ideals may fade. To guard against this possibility, he does two things. He first signs a legal document, which will automatically not give away the land, and which can be revoked only with his wife's consent. He then says to his wife, 'Promise me that, if I ever change my mind, and ask you to revoke this document, you will not consent'. He adds, 'I regard my ideals as essential to me. If I lose these ideals, I want you to think that I cease to exist. I want you to regard your husband then, not as me, the man who asks you for this promise, but only as his corrupted later self. Promise me that you would not do what he asks.'

Participants in the deterioration [improvement] condition were told about some changes that occur many years later:

> Imagine this young man's wife made this promise so the land would [not] go to the peasants. But years later, her husband, now the old Russian, asks her to revoke the document, so as to not give [give] the land to the peasants.

Compared with participants responding to Parfit's original (deterioration) case, those responding to the revised improvement condition agreed more strongly that the old Russian was the same person as the young Russian, free to release his wife from her promise.

Further evidence that direction of change affects identity can be seen in popular examples. Take, for example, Star Trek's 'The Enemy Within'. A transporter malfunction splits a ship's Captain Kirk into two people, one with the properties of Kirk's 'negative side', the other with the properties of Kirk's 'positive side'. Positive-Kirk and Negative-Kirk are both dissimilar from the original Captain, yet the ship's crew refers to Positive-Kirk as 'Captain Kirk' and Negative-Kirk as 'the impostor'. Improved Positive-Kirk is taken as identical and deteriorated Negative-Kirk as non-identical to the original.

Let's take stock. There is a commonplace intuition about the classic (deterioration) Phineas Gage case: it seems (at least to some) that post-accident Phineas is a different person from pre-accident Phineas. The same is true of other cases like Parfit's Russian Nobleman. And there is a commonplace explanation of these intuitions: a certain magnitude of dissimilarity disrupts personal identity.

The experiments and examples discussed here suggest that these classic thought experiments are driven not only by the sheer *magnitude* of change, but also by the direction of change. This tells us something about attributions of personal identity, but there still remains a philosophical question of whether direction of change is actually relevant to the personal identity relation.

There are interesting implications in either case. First, consider if direction of change is *irrelevant* to personal identity. In that case, the finding that direction of change impacts attributions of identity in these seminal thought experiments reveals that such attributions are produced, in part, by an irrelevant factor. This gives a reason to doubt the status of these commonly offered intuitions about seminal thought experiments.

Now consider the other case: we deem direction of change as relevant to personal identity. One challenge for this view is to make sense of some seemingly plausible ways in which the deteriorated individuals still seems to be the same person as the earlier individuals in the vignettes. For example, even if the post-accident man is no longer Phineas, it does seem that he is still the son of pre-accident Gage's mother, that he owns the same house, and that he owes the same taxes. Perhaps someone accepting that the post-accident man is not Phineas might just reject these other judgements; we should conclude from that fact of non-identity that these other intuitions are false. Alternatively, one might allow that certain relations or properties come apart from numerical identity. Perhaps (for example) the post-accident man is non-identical to pre-accident Phineas in some sense, but they are identical to each other in another sense relevant to tax obligations and property ownership. The old Nobleman might be non-identical to the young Nobleman in the sense that he cannot release his wife from her promise, yet he does seem identical to him in the sense that his wife is still his wife.

Personal identity is often taken as a foundation upon which to apply moral and legal notions like responsibility, desert, and blame, but these results suggest the relationship between personal identity and normative notions may be more complex. Personal identity and normativity do share important relations, but these do not flow purely from the foundations of personal identity to normative conclusions: normative considerations exert influence from the start.

Kevin Tobia is a joint doctoral student at Yale University, working in both philosophy and law. This post is based on his paper 'Personal Identity and the Phineas Gage Effect', published in Analysis*. His research is in philosophy, psychology, and law, especially matters of personal identity, emotion, and expertise.*

Image credit: Henrietta Harris, 'Makes Sense'

Monday, 9 May 2016

Related Posts