

Response to Bernard Williams in “The Self and the Future”

SETH CARTER

In the essay, “The Self and the Future,” Bernard Williams presents two instances of a thought experiment that lead the reader to intuit two distinct conclusions on the preservation of personal identity. This becomes particularly notable in consideration of the close methodological similarities between the two cases. Upon analysis of the two experiments and how differences in their presentation may influence our perception of continuity of the self, Williams concludes that the two cases highlight limitations in the methodology of thought experiments that must be recognized before drawing significant philosophical conclusions from them. In contrast to Williams, I propose that while the first thought experiment suffers the theoretical weaknesses he attempts to expose, the second thought experiment provides an insightful exploration into personal identity that is more theoretically sound than Williams gives it credit for.

Sifting through the intuition of both instances of experiment, I argue that the first instance of the thought experiment relies on the intuitive import of psychological features such as memory to conclude more than the proposed experiment actually justifies. Conversely, I contend that the second instance of the thought experiment provides a more accurate philosophical characterization of identity preservation over time. The issues of both thought experiments that Williams highlights can be conceptually resolved by examining the empirical import of the changes proposed in

Seth Carter is a Junior at Indiana University–Bloomington studying economics, philosophy, and mathematics. Upon graduation, he plans to pursue graduate studies in economics. His philosophical interests include ethics, philosophy of economics, and philosophy of religion.

each thought experiment and considering the implications of psychological and physical changes through specific examples. Once this is done, I argue that Williams's thought experiment in its second formulation provides significant support for a theory of personal identity based on the emergence of consciousness from bodily continuity and I consider several potential objections to the claims made throughout for the purpose of rebuttal.

Mapping

In the first instance of Williams' thought experiment, we are asked to entertain a scenario in which a researcher is able to extract and transfer information from two individuals, A and B, such that the memories and character of A and B are interchanged and form the entities body-person A and body-person B, entities that have the originally named bodies of A and B, but with the other person's memory and character traits. Prior to the experiment, a researcher tells both of the individuals that one of their bodies will be given the benefit of a monetary reward, and the other will experience torture after the interchange has occurred. It is then assumed that both would prefer money over being tortured and express this desire. Once the transition is made, and body-people A and B are formed, the researcher arbitrarily decides which entity receives money and which one receives torture. Williams proceeds to explain that whichever one is tortured, because he or she has the memory of a person who asked not to be tortured, will honestly insist that this is not what he or she remembers choosing. Correspondingly, the body-person that is rewarded will be able to honestly state that he or she is receiving the outcome that he or she remembers choosing. Williams concludes from this experiment that the testimony of the individuals provides good reason for thinking that the identity of A has been transferred to a new body, in the form of body-person B, and that the identity of B has been likewise implanted in body-person A. (Williams 182)

Williams entertains a number of further possibilities in the thought experiment to support this claim. An example is that if person A experienced anxiety and person B experienced painful memories prior to the experiment, that they would carry these features with them after the process, such that if body-person A were asked about anxiety, he or she would not remember and instead express disappointment at retaining painful memories. Likewise, body-person B, if asked about his or her painful memories, would similarly not remember such a thing while instead expressing disappointment that his or her anxiety has not lessened. As before, both testimonials appear to confirm a continuity of believed

memory and experience normally characteristic personal identity (Williams 185). From these observations of continuity, the intuition of the thought experiment appears to lead rather smoothly into the conclusion that the personal identities of People A and B have been swapped.

The second instance of Williams' thought experiment asks the reader to imagine that he or she is being experimented upon in the first-person and simply told that he or she will be tortured tomorrow. If the experimenter proceeds to tell the subject that his or her memories and will be taken away before the torture and that new memories belonging to another person will be implanted in the subject's brain, it seems quite clear from the first person perspective that fear about tomorrow's torture will not be lessened, and that in this regard, our impression is that tomorrow's pain will be our pain regardless of the change in information within the brain (Williams 186). Williams then concludes that this concern with our body's future pain indicates that our concern for our future self does not seem to be based on psychological states alone, contrary to the first thought experiment (187).

In order to clarify why the second thought experiment provides a more accurate intuition of personal identity than the first, we first need to consider the self not merely as an indivisible 'monad,' the way that Leibniz or Reid once conceived it, but as the result of an amalgamation of a number of distinct neural and cognitive processes that, in union, provide us with the experience from which our concept of self arises (Reid 109). Although there is certainly practical utility in considering the self as an indivisible entity for purposes of law and emotional simplicity in regards to human relationships, the necessity of breaking apart the components of our experience of self is made necessary by the nature of Williams' thought experiments, which actively address such parts as the information of the brain, memory, character, and our experience of pain piecemeal. By breaking down these conventionally valued aspects of personhood, it is then possible to distinguish between aspects that are merely valued and aspects that are not only valued, but necessary for the appreciation and greater significance of the other aspects of personhood.

I

Problems in the first instance of the thought experiment may be divided into two categories: those that arise from the presumption of more implications than the suggested changes in the experiment actually necessitate, and those that arise from the reliance on the dubious testimony of the hypothetical body-people A and B. I will focus on the former due to

the significance of the observation that many or perhaps even all of these suggested problems are avoided in the second formulation of the thought experiment, which I believe makes it a strong argument for identity based on bodily continuity. Once this is done, the question of the body peoples' sentiment and status as persons may be more carefully broken down.

When examining what actually occurs in the first thought experiment, it may first be asked, is it truly coherent to think an entity can possess the real first-person memories and character once tied to another body by transmitting information alone from one brain to another?

The notion that character follows predominantly from memories or information that may be transferred without fundamentally altering physical composition is itself questionable. Modern neuroscience has discovered numerous correlations between chemical balances whose alteration has been the foundation of treatment for a plethora of illnesses relevant to character, be they diseases of high anxiety, depression, or even schizophrenia (Mitte 141; Kaur and Singh 413). This consideration brings into question the degree to which body-people A and B could respectively have qualitative identity with people B and A, to say nothing of a full assertion of personal identity.

II

Williams could potentially respond to this charge by asserting that mental illnesses do not form the content of personality itself, but rather only frame an existing person's feelings and the direction of his or her established thought processes. To claim this, however, would still be to concede that a large portion of personality in terms of how we act, feel, and take interest in our world around us is not itself the basis of personality, which seems somewhat preposterous, especially when one considers the essential aspects of these features in the context of social relationships where aspects of a person's behavior are valued by others as part of how identity is communicated. Surely, the extent to which depression may cause me to lose my once avid intellectual curiosity in history, prevent me from taking joy in social relationships, and alter my very behavior through emotional and physical exhaustion constitutes a drastic enough change in my personality for me to conclude that the physical composition of my brain plays an essential role in forming the aspects of personality and character that I and my loved ones value. This relation is made even more imposing when one considers that such features have been improved and restored in patients through physical treatments, which would imply causality between physical brain and personality. A more fundamental challenge may even be

presented by the consideration of physical diseases that lead to dementia and memory loss. Although many of us would be hard-pressed to agree that change in a few notable personality characteristics through illness and treatment actually yields a distinctly new personal identity, if these often physically affected personality characteristics are not held as determinative in the first instance of the thought experiment, what necessary condition may then be thought to remain? If none can be found, it would seem that at the very least people A and B are not respectively personally identical to people-bodies B and A. A more substantial argument might be proposed to establish people-bodies B and A as new persons in their own right, but I will leave this question as an alternative topic of discussion.

A further indicator that Williams' first formulation does not coherently accommodate the physical aspects of how brains function may be observed in the experiment's seemingly necessary presumption that psychological states can be thought of as wholly apart from their physical bases to the extent of being transported across space. This may be most directly observed when Williams assumes that the B body-person, who contains the information transferred from person A, would claim to still experience the anxiety he remembers feeling from the time he remembers in body A. In fact, there is no reason to presume that anxiety or numerous other psychological features of a person are dependent on information, let alone information that could conceivably be transferred from one brain to another. A more consistent interpretation of this aspect of the first formulation of the thought experiment would be if body-person B, when asked about the anxiety person A experienced, which he or she has memory of, responds by insisting that his or her anxiety is lesser or even gone, as the presence of such anxiety reducing neurotransmitters as serotonin and adenosine may exist in healthier quantities in the brain of body B (Kaur 416). Although it may or may not be true that anxiety is entirely reducible to physical and chemical properties of the brain, the fact does remain that chemical treatments have an empirical record of reducing the self-reported prevalence of anxiety and its effects (Mitte 141). This physical, rather than information dependent interpretation of changes in personality characteristics could then be extended in the experiment analogously to any number of psychological changes. For example, if body-person B who has the memories of A finds that he or she enjoys ice cream much more than he or she used to due to a greater physical sensitivity to sweetness, or adopts different hobbies due to a greater physical capacity for athletics, the notion that personality is conserved from transferring memory information from one body to another becomes more and more tenuous to consider.

III

Williams' second formulation of the thought experiment in which the subject's fear of future pain is used as an intuitive reason for selfishly considering the same body removed of information as something he or she continues to value as a kind of self is notably immune from the same kinds of inconsistencies present in the first formulation. The key reason for this appears to be that fewer of the features of personal identification of the self are assumed to be predicated on memory and information. Williams or any supporter of the first formulation of the thought experiment as opposed to the second could respond to this charge, however. Why should we think, after all, that a person-body with such distinct memories could be any more be justified to be the same person as A or B by appeal to the entity's personality features better than the new body with the original person's memories and information-based properties of character? This objection is true in establishing that it may not be productive to consider personal identity merely by appeal to how much a personality is preserved. Importantly, however, I do not think that analysis of the self in terms of every one of its parts is ultimately the means by which to establish continuity of identity. Rather, I think that the approach best serves as a methodological means to assess the implicit conditions of the two formulations of the thought experiment with respect to what is known about the functioning of the brain and its implications.

IV

The issue of whether continuity of memory or body as a basis for identity is better supported by Williams' thought experiment, in my mind, seems best characterized as a phenomenological question that emphasizes continuity of conscious awareness across the changes that the thought experiment provides. In this regard, memory and information do not appear to have any necessary direct alteration on the conscious awareness of sensation or the numerical sameness of the mechanism of cognitive reflection itself. In Williams' response to objectors who may consider the thought experiments borderline cases that are conceptually undecidable, an important possibility is neglected entirely: that each person-body A and B in the thought experiment is only a respective reduction of persons A and B whereby the physical framework and cognitive processes for numerical identity are preserved, and yet socially and personally important relevant information removed to be or not be again replaced. The socially relevant facts as well as the content of the memories the body reflects on

are altered, but there isn't a reason to think that the numerical experience of consciousness and phenomenon of cognitive reflection on memories as a physical mechanism are fundamentally altered by changes in the information content of memories. Although there is no doubt room for considerable disagreement on exactly the relation between consciousness and the physical processes of the brain, it seems reasonable to presume that a functioning brain that only undergoes changes in its content of information and memory will maintain the numerical phenomenon of consciousness that may be said to emerge from or consist of the particular brain's physical functioning. The example of an individual who suffers permanent memory loss appears to demonstrate that change in memory or information content alone does not give us sufficient cause to postulate that the neural mechanism by which consciousness arises is that of a distinct consciousness in the way that a wholly independent person experiences a separate consciousness.

Conclusion

Admittedly, the second formulation of Williams' thought experiment does not resolve every theoretical challenge to the notion that consciousness arising from bodily continuity is the most sound, necessary, and sufficient determinant for personal identity preservation. However, in spite of these challenges, thought experiments such as the second that Williams provides in his essay nonetheless offer strong organization of what may be considered either possible or precluded consequences based on some of our best science. Where I believe Williams does not go far enough is in his analysis of the second experiment and in his drawing of hasty conclusions on the limitations of what the experiment can reveal given the insights of modern science.

Works Cited

- Kaur, Sandeep, and Singh, Rajmeet. "Role of Different Neurotransmitters in Anxiety: A Systemic Review." *International Journal of Pharmaceutical Sciences and Research*, vol. 8, no. 2, 2017, pp. 411-416.
- Mitte, Kirsten. "A Meta-Analytic Review of the Efficacy of Drug Treatment in Generalized Anxiety Disorder." *Journal of Clinical Psychopharmacology*, vol. 25, no. 2, 2005, pp. 141-150.
- Reid, Thomas "Of Mr. Locke's Account of Personal Identity." *Personal Identity*, edited by John Perry, University of California Press, 1975, pp. 113-118.
- Williams, Bernard. "The Self and the Future." *The Philosophical Review*, vol. 79, no. 2, 1970, pp. 161-187.