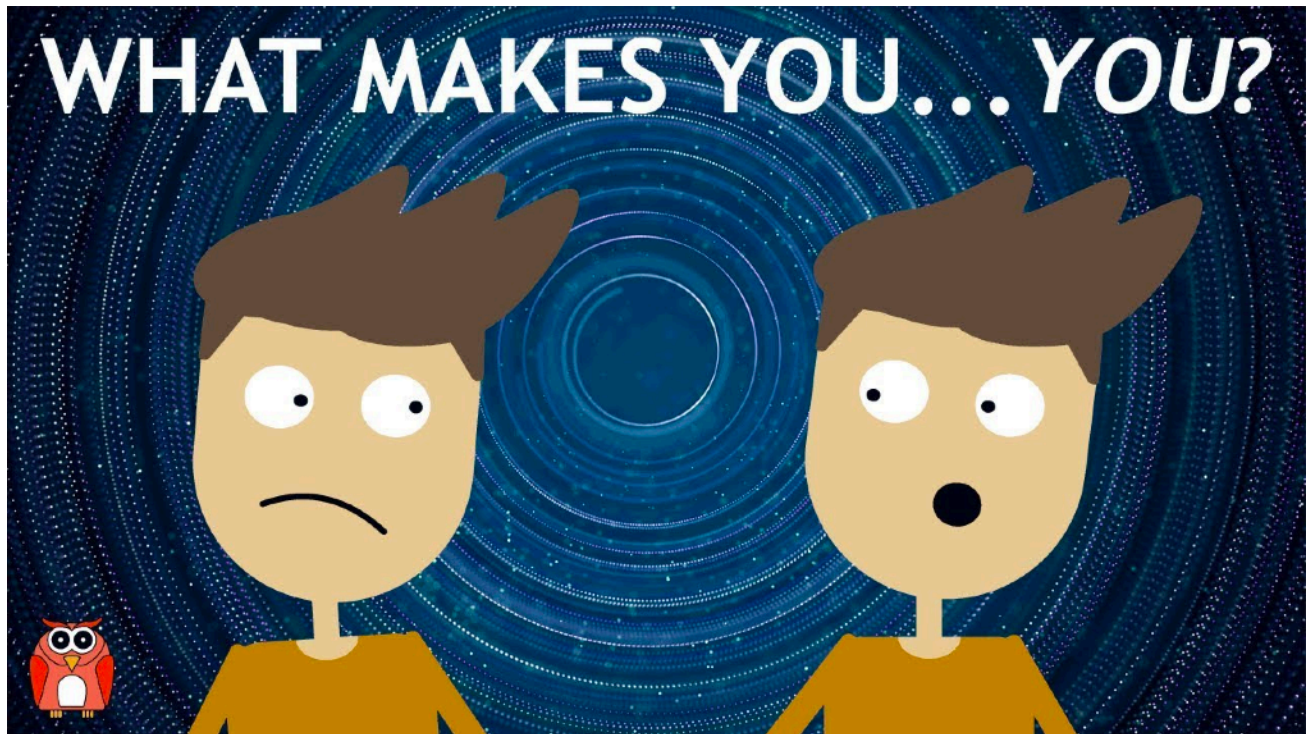


# Hume on Personal Identity: A Critique



**And debating its tenability**

Abhinav Gupta

Monsoon 2019

“Who am I” or “What does it mean to be a person” are common philosophical questions that would arise if we were to contemplate our existence. The problem of ‘personal identity’ as it is so popularly called, has been one of the raging debates in Philosophy. Prior to David Hume, it was mostly assumed that people do have an identity and the idea that there may be no such thing as personal identity was not considered.

David Hume was one of the first philosophers to reject the whole notion of a permanent self as a fiction. Staying true to his empiricist nature, he set out to express his epistemological scepticism by doubting the very existence of a permanent self<sup>1</sup>. Hume says that all we have access to are our impressions and ideas and that all of our knowledge is derived from experience. Everything we know and understand has to be experienced and cannot be ‘innate’.

Hume says that we never constantly experience anything called the ‘self’. Our personal experiences constitute only a series of impressions and sensations. **All we experience is a continuous flow of perceptions**, that take place one after the other. We cannot observe our ‘selves’ and what we truly are and hence, we cannot directly comprehend our identity. We are only aware of what we are experiencing at a given instant of time. Hence, there is no ‘self’, simply because it cannot be perceived and experienced by us.

So there is no such thing as a ‘self’ and even if there is, we are unable to conceive of it precisely **because it cannot be perceived**. According to Hume, all ideas are faint copies of impressions, and in order to have an idea of a ‘self’, this idea must necessarily be derived from an impression. But **there is no impression of a ‘self’** to begin with and hence, the notion of a ‘permanent self’ is nothing but a fiction. For all of our knowledge is derived from experience and even though we experience a large array of impressions, there exists no impression of the ‘self’ that ties all our other experiences together.

We are never aware of our ‘selves’, rather only of what we experience. What we experience is only a continuous flow of perceptions, which occur in rapid

---

<sup>1</sup> David Hume, An Enquiry concerning Human Understanding: <http://web.mnstate.edu/gracyk/courses/web%20publishing/TreatiseLiv.vi.htm>

succession. As Hume himself puts it, “I never catch myself at any time without a perception” and “(I) never can observe anything but the perception”. Yet, we do have a notion of ‘self’. We do think of our ‘selves’ as stable entities existing over time. Hume says that this idea of ‘self’ is just our **natural habit of associating things**. We have a tendency to call something ‘existential’ if the events appear in constant conjunction. We think an object at one time is identical with an object at another time, because there is very little difference. We just see a passing show of appearances, where one follows the other. We naturally make a connection. It is just habit that causally connects the chain of perceptions, giving rise to the idea of ‘self’.

Hume thus **rejects the notion of personal identity** altogether. He calls ‘himself’ a ‘bundle of different perceptions’, and says that the idea of a permanent self is a fiction arising from our habit of associating these perceptions. The self is just a bundle of perceptions, and although connections between these perceptions can be traced back in time with memory, there is no logical evidence of associating them.

Hume’s attack on the notion of personal identity brought out many objections from major philosophers. Penelhum criticises Hume<sup>2</sup> by stating that his argument is an immense blunder. According to Penelhum, an object for Hume cannot have multiple parts. For instance, a piece of music would have multiple musical notes played one after the other in succession. The melody consists of multiple notes, but is still ‘one’ melody. The very existence of the melody lies in its passing away with time. The individual notes occurring one after the other give rise to that ‘one’ melody. Same is the case with the ‘self’; all our impressions and memories are spread out over time, and these multiple parts, much like the musical notes, constitute the ‘self’, just like that ‘one’ piece of music.

The melody analogy put forth by Penelhum does seem to be very convincing, because that would imply that Hume wouldn’t consider an object to have a multiple succession of parts. However, Penelhum misses out on something

---

<sup>2</sup> Penelhum, Terence. “Hume on Personal Identity.” *The Philosophical Review*, vol. 64, no. 4, 1955, pp. 571–589. *JSTOR*, [www.jstor.org/stable/2182635](http://www.jstor.org/stable/2182635).

paramount, something which David Giles very eloquently rebuts<sup>3</sup> in his paper about the ‘no-self theory’. He says that the kind of ‘objects’ that Penelhum is talking about are significantly different from those that are mentioned by Hume. Yes, a melody is a temporal sequence of musical notes and the existence of that ‘one’ melody is necessarily spread out over time. However, the same cannot be said about trees, ships or humans. One could easily fathom an **instantaneous existence** of a ship, or a person, who only exists at a particular instant of time. A tree’s existence need not necessarily be spread out in time. It is logically conceivable to think of ‘self’ to be instantaneously existing. But the same cannot be said about a melody, because at a particular instant of time, a melody is just the individual musical note.

The kind of objects Penelhum talks about are by definition, a temporal sequence of objects and hence, different from the objects Hume concerns himself with. Giles says that a person can momentarily exist and hence we can ask ourselves if a person at a particular instant of time is the same as a person at a prior instant. But there can be **no momentary existence for melodies** because they, by their very definition, are temporally spread out. Thus, the analogy between a unity of a piece of music and existence of self does not hold, simply because self is not a temporal object; it can exist in its entirety in an instant.

Another very common objection to Hume’s attack on the permanent self, which is put forth by Penelhum as well, is that Hume **mixes up numerical and specific identity**. Penelhum’s stance is that in order for us **to posit change, there must be numerical identity**, that is, the object must remain the same through time in the numerical sense. Even if the object’s appearance changes, it would still remain the same so long as it doesn’t change its numerical identity.

Penelhum says that we can see something changing only if we have that something that changes. The object, by definition, must be an unchanging thing, for it to undergo change. For instance, if we were to think of a building that was coloured white initially, and then was painted red at a later instant of time. The two buildings are not the same in the specific sense, but it is the same building

---

<sup>3</sup> Giles, James. “The No-Self Theory: Hume, Buddhism, and Personal Identity.” *Philosophy East and West*, vol. 43, no. 2, 1993, pp. 175–200. *JSTOR*, [www.jstor.org/stable/1399612](http://www.jstor.org/stable/1399612).

that changes. The two buildings are numerically the same because only the colour changes. Similarly, people change with time. If a person becomes taller, numerical identity is presupposed because by gaining height, he is not a different person. He is the same person, but taller. A ten year old boy certainly does not look the same as he was as an infant. The boy has changed from yesterday which presupposes that there is a 'self' that is changing. There is a **numerically identical 'self' as the background against which we can detect change**. And this precisely, is Penelhum's argument at its core. We can only see change against a unity in background. Change can only be observed against something that itself is **unchanging**. Hume is confusing change with numerical difference. But we can only recognise change given the continuity of something that persists for its property to change and for us to perceive that change.

Once again, Penelhum puts forth a very persuasive argument against the no-self theory, but Giles is able to defend Hume with a very simple, yet elegant instance of Hume's ship example. Giles considers a ship whose constituent parts have been replaced over time such that eventually, each nut and bolt used to hold the ship together is completely different from the initial ones. None of the parts of the original ship remain, and yet the two ships look exactly the same. The two ships are specifically identical because of the resemblance. However, **they are numerically different**, as all their parts are different from the original ones. All of its constituents have changed and hence, it is no longer numerically identical.

But even though they're not numerically identical, it would be foolish to say that the ships have not changed! There is an evident change, without numerical identity and hence, serves as a counter argument to Penelhum's stance that to comprehend change, one must necessarily posit numerical identity. For we can certainly imagine a situation where we replace every bolt and change it numerically, and yet cause a change. The two ships are indubitably not the same. Change has occurred but **there is no numerical identity underlying that change**. Hence, we can have instances of change which do not presuppose numeric identity.

Another counter to Penelhum, along the lines of the ship example, could be that of **cloning**. There is no numeric identity amongst the clones and yet, one can perceive a total resemblance. According to Penelhum, there has to be an underlying thing against which change is seen. But the clones are not numerically similar, and yet a change has taken place. Giles shows that we can have an identity in each specific moment without positing any form of continuity. We can exist in each moment of time with all our memories. The next moment is a different person, but with the same memories.

Thomas Reid also objected<sup>4</sup> to Hume's attack on personal identity and brought out the question concerning moral responsibility. Reid says that we hold a person responsible for actions that he committed in the past. But according to Hume, the person would not be the same as the one who committed the action in the past and hence, would not be responsible for his actions. So a person could not be punished for a crime that he may have committed in the past, because he is not the same as the person in the past, thereby demolishing the very idea of moral responsibility.

Lasse Nielsen defends Hume<sup>5</sup> by saying that this link between past crime and present judgement is of a 'hereditary' nature. The person in the present is certainly not identical with his 'self' in the past, but he is the **result of his past actions**. In Nielsen's words, "if I am being punished for a crime, I may not be the same person as the person who committed the crime, but because I am a result of that person, his responsibilities are now my responsibilities." He says that there are no real connections between past and present selves, but for moral judgement to hold, we create a fictional hereditary link to hold people responsible for past actions, thereby countering Reid's objection.

David Giles and Lasse Nielsen thus defend Hume against some of the most popular objections. But is there any tenable rebuttal to Hume's notion of a fictional permanent self? Hume says that it is through association that we get the notion of the permanent self. But how is this association itself possible? How

---

<sup>4</sup> Robinson, Daniel N. & Beauchamp, Tom L. (1978). Personal Identity: Reid's Answer to Hume. <https://www.jstor.org/stable/pdf/27902536.pdf>

<sup>5</sup> Nielsen, L.: Defending Hume's Theory of Personal Identity and Discarding the Appendix. In: *Ostium*, roč. 12, 2016, č. 2. <http://ostium.sk/language/sk/defending-humes-theory-of-personal-identity-and-discarding-the-appendix/>

can we have a stable habit if we are spread out in time with no connection? This is exactly what Immanuel Kant attempts to answer.

Prior to Kant, one would generally associate apriori knowledge with analytic judgements, and aposteriori with synthetic judgements. Hume put knowledge into these two categories. Kant put forth a third type of knowledge by associating synthetic with apriori. He said certain principles are synthetic apriori knowledge which are necessary truths known apriori that tell us something about the world. We need these apriori structures of thought in order to have meaningful experiences. Kant calls this '**transcendental**' because it is not realisable in experience. Kant does not deny empirical observations, but he says that we also need these transcendental ideal structures to enable us to make sense of the world.

Hume says that we associate things by habit. But what really is the reason for this constancy? How does he then bring in memories and habits? For habits presuppose a stability and someone needs to have **experienced things continuously in order to form a habit**. It is through association with our past selves that we get the sense of a temporal continuity. Even the fiction of a permanent self wouldn't be possible if it weren't for the notion of time. Temporality is essential for experience.

Revisiting Giles' modified ship example, he says that even though both the ships are not numerically identical, change has occurred as the parts have changed. But who is perceiving this change? One would probably not even be able to distinguish between the two, if it was not known prior that the nuts and bolts had been changed. Had the person not known the previous ship existed, he would have thought that it was a new ship and the problem itself would not have arisen. The problem arises only because of the experience of an earlier ship! Hume's empiricism **presupposes a continuity of experience** to perceive identity, and this identity does not lie in the ship itself, but in the witness who is perceiving this transformation.

On similar grounds, how does one perceive is something is a clone of an object? We ought to have perceived the original object, in order to be able to comprehend its clone. If we perceive both the object and its clone at the same

time, we perceive the numerical distinctness. But if the object disappears and its clone comes into being, which is exactly like the original object, we wouldn't know the resemblance without having experienced the previous object, which now no longer exists. This shows that there must be a continuity of experience in order to perceive resemblance. And this resemblance does not lie in the ship itself. **It lies in the person witnessing the transition!** This is precisely what Penelhum ought to have said to counter Giles' argument successfully.

Kant's account of cognition is thus able to successfully refute Hume's notions of a fictional permanent self and hold up Penelhum's argument against Giles as well. Hume presupposes constancy of association in his empiricist frame. But **in order to make these associations, there must be a permanent self.** The unity of self must be assumed in order to experience things around us. We have to reproduce the past in the present for which unity of the self is essential. There has to be a continuity in experience in order to be able to perceive likeness. This cannot lie in nature because we do not have access to nature in itself. It lies not in the object, but rather in the subject who is witnessing and perceiving. Kant thus, makes the **transcendental move** to explain the continuity of experience and is able to defeat Hume's no-self theory.