

HUME AND SELF-IDENTITY

HUME Y LA IDENTIDAD PERSONAL

CARLOS EMILIO GARCÍA DUQUE

Universidad de Caldas, Colombia. carlos.garcia_d@ucaldas.edu.co

RECIBIDO EL 20 DE ENERO DE 2009 Y APROBADO EL 17 DE ABRIL 2009

RESUMEN ABSTRACT

En este artículo deseo examinar las tesis de Hume acerca de la identidad personal. Pese a ciertas opiniones negativas sobre estas tesis, trataré de ver como se pueden aprovechar y de qué modo la perspectiva de Hume guarda correspondencia con nuestras intuiciones al respecto. Para tal fin, replantaré el argumento de Hume sobre la identidad personal. En la segunda parte del artículo, emplearé sus puntos de vista para responder a ciertas cuestiones enigmáticas sobre casos de doble personalidad. En la sección de cierre haré algunos comentarios sobre la acusación de Penelhum, según la cual el filósofo escocés cometió un error en investigación conceptual.

In this paper, I examine Hume's views on personal identity. Despite some negative opinions on his theory, I will try to see how to make something useful from it and to explain the way in which Hume's final views on this matter are consistent with our intuitions. To this end, I will recast Hume's argument on personal identity. In the second section of the paper, I will use his views to answer some puzzling claims about split personality. In the concluding section, I will advance some comments on Penelhum's charge that the Scottish philosopher committed a mistake in conceptual investigation.

PALABRAS CLAVE KEY WORDS

Doble personalidad, Hume, identidad personal, investigación conceptual, Penelhum.

Conceptual Investigation, Hume, Penelhum, personal identity, split personality.

It is hard to think of a more controversial topic than Hume's views on personal identity. The discussion he gives to this matter is, without doubt, incisive and penetrating. Yet, his overall theory has been considered by many to be discouraging and mistaken. As we learn from the Appendix to the *Treatise*, Hume himself was not completely satisfied with what he had achieved on this labyrinthical topic. But some commentators have taken his own remarks to the extreme of considering Hume's account of personal identity as an example of a complete philosophical failure. This is, of course, not the kind of verdict that could generate widespread agreement. On the contrary, there are many philosophers who think that much can be made of such account. Even accepting that Hume's treatment of personal identity is affected by serious mistakes, they are willing to recognize that his final picture is perfectly consistent with our intuitions. It will be my primary objective here to recount Hume's argument on personal identity, so I will devote the longest section of the paper to this goal. In the second section, I would attempt to use his views to answer some puzzling claims about split personality. In the last section of my paper, I will advance some comments on Penelhum's charge that the Scottish philosopher committed a mistake in conceptual investigation.

HUME'S ACCOUNT OF PERSONAL IDENTITY

In the opening paragraph of Section V (*Treatise*, Book I) Hume characterizes what he takes to be the traditional view on personal identity. According to such view, we are always intimately conscious of what has been called "the self". Furthermore, both dyachronic and synchronic existence of the *self* is open to sense experience, and because of our continuous experience of it, we become certain "*beyond the evidence of demonstration*" of its perfect identity and simplicity. The traditional view even holds that strong sensations and violent passions instead of hindering our perception of the self, fix our awareness of it more intensely. Unfortunately, Hume thinks, this whole view is completely mistaken. And the source of the error is not a minor misunderstanding. Contrariwise, it stems from the very nature of the line of reasoning which leads to the conclusion that we discover an enduring and identical *self* just by reflection.

Recall that in Hume's philosophy an idea of the *self* as that pictured in the traditional view has to be traceable to its corresponding impression. So he considers himself entitled to ask "*from what impression is this idea of the self derived?*", a question that strikes him as impossible to answer without falling into contradiction, but that anyway has to be resolved if we are to make sense of this notion of *self* at all. The problem is that the *self* is precisely that to which impressions and ideas are supposed to have a reference. Introspection can show that the traditional view seems to be inadequate, because instead of yielding the idea of a continuously existent entity that remains invariable the same throughout the whole course of our life, it produces the notion of change, diversity and succession of many passions and sensations that do not coexist ever. Hume thinks that since it is impossible to locate the suitable impression which could give rise to the idea of the *self*, we must conclude that there is no such an idea. As for our idea of the existence of whatever is being designated by the word "self", it seems to be triggered by the existence of all sorts of impressions and feelings:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* without a perception, and never can observe anything but the perception. When my perceptions are remov'd for any time, as by sound sleep; so long I am sensible of *myself*, and may truly be said not to exist. And were all my perceptions remov'd by death, and cou'd I neither think, nor feel, nor see, nor love, nor hate after the dissolution of my body, I shou'd be entirely annihilated, nor do I conceive what is farther requisite to make a perfect non-entity (Hume, 1972: 84).

This is certainly a remarkable view to hold. The *self* (perhaps we can equate it to the mind) does not have a continuous existence. Hume literally means that the *self* ceases to exist in periods like sound sleep, total lack of perceptions (ideas) and death. I cannot explore completely the implications of this view for the purposes of adjudicating about so controversial topics as the eternal life that some religions attribute to the individual *self*. However, refusing to grant existence to the *self* during periods of deep sleep or full empty-mindedness (were such phenomena

realizable) seems to be a thesis which many would find at least surprising. On the other hand, it is necessary to grant that Hume's initial rejection of the idea of the *self* seems to be well supported by the contention that allows only intermittent existence to it. Moreover, he is so confident on the adequacy of his comments above that quickly gives up any hope of meaningful exchange on this matter with persons who claim to have (along the lines of the traditional view) a notion of their own selves as enduring entities.

But setting aside some metaphysicians of this kind, I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement (*Ibid.*: 84-85).

This famous passage has been the object of intense analysis. After many disparate interpretations it is clear (at least) that Hume wishes to emphasize the changing character of the representation that many impressions, so to speak, perform in the theater of the mind. We do not have, therefore, an idea of our selves that is different from the particular impressions and ideas that we entertain in our minds. But this rhetorical talk about “we” could be misleading. As Biro has pointed it out, Hume wants to respond to the ontological question: what is the *self*? By laying down the way to the related problems of the nature of my belief that I am a *self*, and of the source from which such belief comes. Hume's strategy, asking the questions from the first-person point of view here, renders irrelevant “*the easy answer that to believe in one's identity is to believe that one is, or is at least associated with, an enduring body*”¹.

Now, if a *self* is just a bundle of different perceptions, why do we succumb so easily to the temptation of forming the idea of an enduring self? And why do we attribute to it (mistakenly) the notion of sameness? Hume thinks that the source of this error lies in the resemblance amongst what one might want to call “*different time slices of the mental realm*”. Notice

¹ Biro, J. “Hume's new science of the mind”. In: David F. Norton (Ed) *The Cambridge Companion to Hume*. New York: Cambridge University Press. 1993, p. 47. It is noteworthy that Hume neglects any possibility of solving the problem of personal identity by considering attributions of identity to the body. And I believe this is important because for many ordinary purposes identity is decided precisely based upon physical appearance and the resemblance of past and present images.

that one is directly aware of one's mental states, there is not an internal observer (different from oneself) who contemplates what is going on in the mind. But once I obtain knowledge of my *self*, I have epistemic access of another mental object whose relation with the rest needs to be explained. It is in the course of providing this explanation that we come to contrast the notions of sameness and diversity, since, by doing so, we conceive of an object that is capable of remaining invariable through time as well as of different objects that exist in succession.

In the passages that follow this discussion, Hume explains the nature of the error and begins to develop his own theory of personal identity. There are cases, he argues, where we make attributions of identity to something, in spite of the fact that this involves coupling together a set of different things or impressions. His examples are very common in the literature about identity and material constitution. Cases of alteration of a piece of matter, the replacement of the parts of a ship, the reconstruction of a church and the notorious changes that take place in plants and animals. Hume wants to read in all of these examples the instantiation of a general thesis that would explicate our tendency to make the attributions of identity: what all these cases have in common is that they consist of a succession of related objects. As to the error of saying that *different* (but related) objects are the *same*, it seems to be not only inevitable but conveniently sanctioned by linguistic uses. We fall into it even before being aware that a mistake is being made; once we realize the situation we might correct some attributions, if the conversational context makes that advisable, and tolerate the rest by the force of habit².

Hume distinguishes two notions of identity. He calls them "perfect" and "imperfect" identity and we could assimilate them to numerical identity and specific (ordinary) identity³, respectively. He also believes that we often attribute identity in an improper sense, like when we attach the label to variable or interrupted objects. Sometimes, when we become

² "In order to justify to ourselves this absurdity, we often feign some new and unintelligible principle that commits the objects together, and prevents their interruption or variation. Thus we feign the continu'd existence of the perception of our senses, to remove the interruption; and run into the notion of a soul, and self, and substance to disguise the variation" (Hume, 1972: 86).

³ I would have preferred to use for this sense of identity the word "similarity" or perhaps the notion of kind membership. However explaining this terminological point would have taken me very far away from my main objective here.

conscious of a break in the continuity of our thought, we withdraw the attribution of identity, but if this not the case we leave things like that. It is in this sort of cases when we worsen things by postulating a “fiction” that would support the mistaken attribution. When it relates to human minds, the fiction is no other than the *self*.

Let us turn now to the specifics of personal identity. As in the case of sameness of objects in general, neither does the idea of personal identity belong to the different perceptions that we encounter in the mind, nor it unites them together. It is something that we attribute paying attention to their course in imagination by means of three relations: resemblance, contiguity and causation. Then,

[...] it follows, that our notions of personal identity proceed entirely from the smooth and uninterrupted progress of the thought along a train of connected ideas, according to the principles above-explain'd (*Ibid.*: 88-89).

Hume, then, goes on to explain in detail how the three relations just mentioned can contribute to our attributions of personal identity. He asks us to imagine that we can see into the self of another individual and come to contemplate the succession of perceptions which constitutes his mind. Insofar as those perceptions are mental states they belong to the same kind, this seems to be a natural thing to say. However, we need to make a comparison to discover the similarity so that the principle of resemblance be preserved. For this endeavor we require the help of the memory which not only discovers the identity, by contrasting past and present perceptions, but supports the idea of identity, “*by producing the relation of resemblance among the perceptions*”.

Memory, aided by the natural relation of causation, unites the perceptions that collectively constitute the mind. The presence of memories among my perceptions gives birth to the idea that I am a temporally extended being. In the thought experiment that Hume suggests, the observer notices a resemblance between certain past and present perceptions had by another person (the subject). The resemblance relation holds between the subject’s present recollection and the past perception the subject recalls. Upon noticing this resemblance, the observer comes to

regard the subject's perception as 'the continuance of one object', comes to deem the subject as identical through change⁴. Lastly, the relation of contiguity is necessary to secure the idea of one single thing (the *self*) where we have knowledge of a succession of related perceptions being entertained by the same subject. It is due to the temporal interruptions in the awareness of the activities of the *self* that Hume claims one can properly deny existence to the *self* during deep sleep⁵.

However necessary to the notion of the identity of the self may be the relations of resemblance, contiguity and causality, the leading role in producing the very idea of the *self* is reserved to memory alone. Had we no memory, Hume contends, we would not have ever a notion of causation, nor consequently "*of that chain of causes and effects, which constitute our self or person*". In this respect, memory produces and discovers personal identity, by exhibiting the relation of causality that bears between different perceptions. Memory also allows us to recognize ourselves as one individual over time, simply by extending the same chain of causes that foster the understanding of personal circumstances and particular stories that we might have forgotten about⁶.

After establishing that the mind can be conceived as a "*system of perceptions link'd together by the relation of cause and effect*", Hume gives his famous political analogy:

⁴ Bricke describes the experiment in the following way: "*basically, a distinction must be drawn between the self as subject and the self as (self-) observer. More precisely, a distinction must be drawn between those resembling perceptions that provide the basis for one's judgement of self-identity and those perceptions that in some way constitute one's awareness of the former perceptions and of their resemblance to one another*". Bricke, J. "Hume on Self-Identity Memory and Causality". In: G. P. Morice (Ed.) *David Hume. Bicentenary Papers*. Austin: University of Texas Press. 1977, p. 169.

⁵ Hume's remark about everyone's inability to recall exactly the thoughts and impressions had in a specific and remote moment of our past stories is very illustrative of what he means by the discontinuity of the self.

⁶ For a thorough treatment of the problem of memory see: Biro, J. "Hume on Self-Identity and Memory". In: *The Review of Metaphysics*. XXX. p.117. 1976.

In this respect, I cannot compare the soul more properly to anything than to a republic or commonwealth, in which the several members are united by the reciprocal ties of government and subordination, and give rise to other persons, who propagate the same republic in the incessant changes of its parts. And as the same individual republic may not only change its members, but also its laws and constitutions; in like manner the same person may vary his character and disposition, as well as his impressions and ideas, without losing his identity. Whatever changes he endures, his several parts are still connected by the relation of causation (*Ibid.*: 90).

Hume thinks that the same republic can persist despite the unceasing change of the citizens which make it up because there is some very general pattern of organization that is preserved even if the laws or the constitution of such political entity are amended. Is there a limit to the extent of the reorganization that we want to allow in a political entity without feeling constrained to declare its identity lost? I think the answer must be positive, as it is well illustrated by the historical examples of great revolutions that lead people to refer to what was a certain political entity with a different name. This suggests that there are some basic conditions without which we would agree to have large groups of persons inhabiting the same geographical place but not a commonwealth. Likewise, if the analogy works properly, we would want to extend the same result to the case of the human mind and maybe try to find out what the principle of organization would be for it. As far as I can see, the pattern would be provided by nothing else than the principle of causality, responsible for establishing the appropriate connections among the ideas in the mind and sustaining the notion of identity.

I shall not attempt to exhaust the incredibly high number of problematic questions that can be raised to Hume's notion of personal identity. He himself thought that all the nice and subtle questions concerning this topic could never possibly be decided. Let me finish this section by pointing out that no matter how unsatisfactory one might find his account, there are many respects in which is not far from what common sense would dictate to solve the questions Hume was trying to tackle.

Dr. JEKYLL AND Mr. HYDE

We are all familiar with cases of split or multiple personality. The *locus classicus* is a known literary piece by R. L. Stevenson: *Dr. Jekyll and Mr. Hyde*. According to the novel, they are two completely different personalities who share one body. As noticed above, for ordinary purposes one popular criterion of identity for persons relies on physical appearances. The legal system, for example, makes a big deal of witnesses identifying suspects by the way they look. This does not mean that the obvious drawbacks of this method of identification had been overlooked. Although frequently reliable, it is imperfect and deemed to failure under many circumstances. Take just the borderline cases of physical resemblance between siblings or the so-called identical twins, not to mention possible cases of resemblance between persons who are not family related. Here our intuitions vacillate and the legal system has developed more precise ways of establishing identity (e.g. by using tests of DNA).

As Biro notices it, Hume dismisses the quick answer to the problem of personal identity in terms of the association of the *self* with an enduring body. Recall that mankind is just a bundle or collection of different perceptions; each one's own perceptions as they are united by the memory. So, it seems that to answer the question "*Is A the same person as B*" under Hume's account of personal identity, we cannot resort to the physical traits of the body. And this does not seem to be very different from what most refined views will advise. Hume's answer to the question just asked will probably be: "*A and B are the same person iff, they share the same mind*". But on his account minds are individuated by a private collection of perceptions which is united under the notion of one self having awareness of his entertaining those perceptions over time. To put this in simple words, a mind would be constituted by a set of private memories. Let us call the set of memories N. Then, *A* and *B* are the same person if they have the same N.

Let us turn, for a moment, to ordinary life. If Mr. Hyde visited Tilbury Docks at 9:30 p.m on December 18, 1887, then Dr. Jekyll did too, and did whatever Mr. Hyde did in that place. But according to the plot, Dr.

Jekyll and Mr. Hyde are different personalities and they do not necessarily communicate with each other. Now Dr. Jekyll, the man, is a person and he did those things at Tilbury Docks, but he is not the person that did it? How can he be a different person from Mr. Hyde? After all, we are able to see just one man whenever any one of these personalities engage in his favorite activities, goes to some place and does whatever he does. Suppose that they do not ever communicate to each other and that they do not share memories, that is, that Dr. Jekyll is totally unable to recall anything that is part of Mr. Hyde's memory, and *converso modo*, the latter cannot recall anything that the former does; in short, they do not have the same N. This being so, I would venture the hypothesis that they would fail Hume's test of personal identity and had to be considered as two different persons⁷.

Such conclusion however is hard to accept. But the legal system provides many examples in which these kinds of situations are solved in one way or another. Yet the puzzle remains unanswered. I can be punished by my past crimes, but no one can be justly punished by anyone else's crimes. I can fear my future pain, but no one can fear (in the appropriate way) the future pains of others. While others can wish that I had not done certain things, only I can regret my doing so. Accountability was used as a measure of personal identity by Locke, could it be used by Hume? Independently of the answer to this question it seems desirable to reconcile the results of our test of identity with our intuitions for the present case. I suggested before that Hume would probably support the defense attorney's view that Dr. Jekyll cannot be justly punished by what Mr. Hyde did because they were not the *same* person, but perhaps my suggestion is misconceived.

It has been well established that Hume's views on personal identity have to be interpreted as referring to self-identity and that this kind of identity refers to the mind and not to the body. We have seen already the role that his theory attributes to memory, what about personality?

⁷ We would have to give a similar answer for cases of multiple personality under the proviso that the personalities involved do not share memories. But I have personally witnessed a case of a man that engaged in lengthy conversations in which several of his personalities participated and cross-examined each other in ways that suggest they did share memories. In a case like this, the answer has to be different.

Hume did not consider personality as a necessarily stable characteristic of the individual. He was convinced that one might vary completely one's personality without losing one's identity⁸. Thus, that Dr. Jekyll and Mr. Hyde have different personalities is not sufficient to conclude that they are not the same person. It would be interesting to see whether this kind of puzzling cases can be solved by Hume's criteria of personal identity and to determine up to what extent we can reconcile the picture of our ordinary intuitions with the humean solution. I shall not attempt to complete this task here, but I suspect that Hume's views do not work very well with abnormal cases.

IS HUME'S ACCOUNT BASED ON A LINGUISTIC MISTAKE?

In his insightful essay *Hume on personal identity*, Terence Penelhum attributes Hume's failure to solve the problem of personal identity to his mistaken idea that invariance is a necessary condition of numerical identity. Penelhum also argues that Hume's discussion of personal identity is infected by an improper analysis of 'identity' which is his starting point. Recall that Hume defined 'perfect identity' as identity that could only be correctly attributed to what could remain "*invariable and uninterrupted*". When he carries out the project of characterizing a suitable notion of identity for the mind, he is looking precisely for these characteristics. Given this starting point, the project of accounting for mental identity amongst the diversity of perceptions must either i) demand a return to some of the notions of identity that Hume has rejected in the first part of section V or ii) legitimate the fiction that would be generated when we ignore the diversity of our perceptions. Penelhum thinks that the Scottish philosopher would be committed to regarding "*on a stricter view*" all accounts of personal identity as belonging to the class that he has reasons to reject.

On the other hand, Penelhum reads very strong theses in Hume's account of personal identity. He thinks, for instance, that according to Hume we commit an abuse of language every time we use the same word to

⁸ "[I]n like manner the same person may vary his character and disposition, as well as his impressions and ideas, without losing his identity. Whatever changes he endures, his several parts are still connected by the relation of causation" (Hume, 1972: 90).

designate a changing entity, and *a fortiori*, every time we use the same name to call a person⁹. This is perfectly plausible although does not fit very well many passages in which Hume refers to a person's identity without any hesitation or any fear of falling into any language trap. More interesting, however, is Penelhum's charge that Hume conflates, when he arrives at the condition of invariance, numerical and specific identity. The commentator notices (rightly, I believe) that when two things are said to be the same in the numerical sense, they are indeed one thing; and that to remain unchanged normally means to remain the same in the specific sense. But nothing prevents one thing from remaining the same in the numerical sense without doing so in the specific sense. For example, object *x* can be numerically the same despite the changes it undergoes; and there is a sense in which it has to be self-identical if it is to change at all.

We can answer Hume's misgivings about sameness by saying that "*the pairs of expressions, (a) "numerically the same" and "containing many parts" and (b) "numerically the same" and "changed," are not pairs of contradictories*"¹⁰. This means that there is no mistake when one claims that a succession of related objects form a unit of a certain kind, or that one and the same thing can experience multiple, and sometimes radical changes. Furthermore, nothing in an object which remains invariable and uninterrupted (e.g., a continuous sound) requires an attribution of identity and nothing in a succession of different but related ones (e.g., the notes that constitute a musical theme) requires a denial of identity. It is the context, Penelhum argues, what makes one way or the other the natural response.

I find these comments appropriate and correct. I do not see, however, how they are enough to prove that Hume's account of personal identity was caused by a linguistic error. It is clear, from section V of the *Treatise* that Hume was not setting up conditions for correct ascriptions of unity and identity to perceptions. He was trying to explain the source of an

⁹ We are May making a mistake "in referring to a person from day to day as the same person (in using the same proper name, for example), or in referring in this way to anything that has changed in the slightest. For strictly speaking, a changed person will be literally another person". Penelhum, T. "Hume on personal identity". In: Chappel, C. *Hume*. Notre Dame: University of Notre Dame Press. 1968, p. 224.

¹⁰ *Ibid.* p. 228.

error: how philosophers come to form the fiction of the *self*. Moreover, Hume attributed the error of ascribing unity and identity to things which do not satisfy his criteria, to their misleading resemblance to things that do. This could explain Hume's insistence on attributing perfect identity only to things that are uninterrupted and unchanged. But, it seems to me, setting aside the more problematic cases of the interruption of the self while deep sleep, Hume's overall account of personal identity is not very far from what our intuitions might support. I regret not having the time to give a better treatment of Penelhum's wonderful essay on Hume. However, I suspect, he is too eager of making his case for the thesis of Hume's committing a mistake in conceptual investigation, that he does not realize he can be guilty of the same crime.

REFERENCES

- BIRO, J. (1976). "Hume on self-identity and memory". *The Review of Metaphysics*. XXX. p. 117. Washington: Philosophy Education Society of America.
- _____. (1993). "Hume's new science of the mind". In: NORTON, D. F. (Ed) *The Cambridge Companion to Hume*. New York: Cambridge University Press. p. 47.
- BRICKE, J. (1977). "Hume on self-identity memory and causality". In: MORICE, G. P. (Ed.) *David Hume. Bicentenary papers*. Austin: University of Texas Press. p. 169.
- HUME, D. (1972). *A Treatise of Human Nature*. Edited by Charles W. Hendel. New York: Charles Scribners's Sons.
- PENELHUM, T. (1968). "Hume on personal identity". In: CHAPPEL, C. *Hume*. Notre Dame: University of Notre Dame Press. p. 224.