

## HUME'S SECOND THOUGHTS ON PERSONAL IDENTITY\*

**Sunny Yang**

Minerva Liberal Art College  
Hankuk University of Foreign Studies  
Oedae-ro, Mohyeon 81, Choingu  
Yongin, Gyeonggi, 17035, Korea(South)  
E-mail: [sunnyyang@hufs.ac.kr](mailto:sunnyyang@hufs.ac.kr),  
[sunnyyang@hotmail.com](mailto:sunnyyang@hotmail.com)

**Abstract.** *In this paper, I present an interpretation on how Hume can escape from his intellectual ordeal concerning personal identity in the Appendix of the Treatise. First of all, I present the source of Hume's despair to offer an interpretation on what would have truly bothered Hume in the Appendix, and I identify several lines of interpretation. Recently Jonathan Ellis has distinguished various ways of understanding Hume's predicament. Of the four groups of explanations that Ellis distinguishes, in this paper I elaborate on the three that Ellis does not sufficiently explicate, addressing some key issues that Ellis missed. Last, I offer an alternative reading of Hume's difficulty, based on Dennett's ideas on the matter, and make a suggestion about what Hume ought to have said about these problems.*

**Keywords:** *Personal Identity, Hume's second thoughts, connecting principles, Hume's inconsistency, homunculi model*

The section of *A Treatise of Human Nature* entitled "Of personal identity" in Book 1 has attracted much attention, among other things, because Hume himself expressed his dissatisfaction with it, saying, in the Appendix:

Upon a more strict review of the section concerning *personal identity*, I find myself involv'd in such a labyrinth, that, I must confess, I neither know how to correct my former opinions, nor how to render them consistent. (T633)<sup>1</sup>

Then he presents his difficulties in these words:

having thus loosen'd all our particular perceptions, when I proceed to explain the principle of connexion, which binds them together, and makes us attribute to them a real simplicity and identity; I am sensible, that my account is very defective, and that nothing but the seeming evidence of the precedent reasonings cou'd have induc'd me to receive it. (T635)

---

\* This work was supported by Hankuk University of Foreign Studies Research Fund of 2018.

I expressed some ideas of this paper in my PhD Thesis defended at Durham University in 2006. I am thankful to anonymous referees of *Problemos* for invaluable suggestions. I also thank to those audiences at the workshop on Hume and Naturalism in Durham University 2015 for their valuable discussions of previous version of this paper. Especially, I would like to thank Saul Traiger, Andy Hamilton, Peter Kail and Ben Smith for their comments.

I want to extend my thanks to Jesus Illundáin-Agurruza for his comments in personal correspondence.

---

<sup>1</sup> Hereafter, all references to Hume's *Treatise* will be cited with a "T" and page number or section.

After these passages Hume says:

All my hopes vanish, when I come to explain the principles, that unite our successive perceptions in our thought or consciousness. I cannot discover any theory, which gives me satisfaction on this head.

In short, there are two principles which I cannot render consistent; nor is it in my power to renounce either of them, viz. *that all our distinct perceptions are distinct existences, and that the mind never perceives any real connection among distinct existences.* (T635-636)

He then says that

Did our perceptions either inhere in something simple and individual, or did the mind perceive some real connexion among them, there would be no difficulty in the case. (T636)

In the Appendix passages from above (T635-636), Hume seems to claim that his theory of personal identity involves unavoidable inconsistencies. Many commentators have discussed the problem that Hume is complaining about to ease him out of his difficulties, because finding unavoidable inconsistencies in one's theory is definitely ill-advised. However, there has been no general agreement as to what leads Hume to confess his difficulties and retract his theory of personal identity. The consensus among commentators so far seems to be that the problem that Hume found in the Appendix is his earlier account of personal identity (T 1.4.6) rather than a new one, about which, upon review, he found reasons for rejecting it. Many commentators have offered interesting suggestions about what the problem in his earlier account may be. In doing so, they raise important questions about Hume's philosophical system. One point on which all seem to agree, howev-

er, is that Hume's *two principles*<sup>2</sup> that he refers to are logically compatible and are the grounds for his own skeptical accounts of the self, external objects, and causation. The supposedly inconsistent principles are not inconsistent with one another, but rather with some other *third principle* which he never explicitly identifies. What this third principle might be is controversial. In this paper I examine several interpretations of what this principle might be, in the hope of providing a clue as to how to resolve the alleged inconsistency. I will finally consider alternative ways to ease Hume out of his difficulties. In order to do this, we need to consider several interpretations concerning the source of Hume's labyrinthine difficulty in the Appendix.

## 1. The Source of Hume's Despair

### 1-1. The Problem of individuation and ownership

Recently Jonathan Ellis (2006) has distinguished various types of interpretations of what Hume came to eventually realize in the Appendix (T633-636), "upon a more strict review of the section concerning *personal identity*". Ellis divides the explanations into four groups.

According to Ellis's classification, the first group sees Hume's difficulty as the problem of ownership of perceptions: "what actually unites our successive perceptions into one mind or consciousness" or "what actually unites them together to make up

---

<sup>2</sup> The supposedly inconsistent principles that Hume mentioned in the Appendix (T636) are "*all our distinct perceptions are distinct existences*", and "*that the mind never perceives any real connection among distinct existences*".

one mind”.<sup>3</sup> I shall dub this group of interpretations “the problem of individuation and ownership”. Don Garrett provides an ingenious variation on this theme. According to Garrett (1981: 350), the most fundamental difficulty with Hume’s account and the most likely cause of his having second thoughts, is that he would assent to each of the following jointly inconsistent propositions:

- (A) All our distinct perceptions are distinct existences.
- (B) The mind never perceives any real connection among distinct existences.
- (C) The ownership of any perception is determined either by its causal relations and its relations of resemblance, or by its perceived real connection to (inherence in) a distinct substantial self.
- (D) The causal roles of qualitatively identical objects (distinct existences) can differ only in virtue of differences in spatial or temporal location, unless the mind perceives a real connection (a necessary causal connection) between at least one of them and some other distinct object.
- (E) Many kinds of perceptions are “no where”.
- (F) It is possible that two qualitatively identical perceptions of any kind, including those that are “no where”, should occur in different minds at the same time.

With this long list of propositions, Garrett tries to pin down the heart of Hume’s problem. Now it seems that D, E, and F are the core of Garrett’s argument. He supposes there are two spatially non-locatable perceptions (for example, sounds, tastes, smells and even passions) in the minds of A and B respectively and they may occur simultaneously.<sup>4</sup> Then he asks: how, on Hume’s theory, are we to assign these two perceptions to different minds? He thinks that Hume would answer: not by resemblance, for we have assumed that they are qualitatively identical. If resemblance could not be a candidate, the only alternative would be causality. But according to Garrett, that doesn’t seem to work either, since distinct causes can be separated only on the basis of spatial relations, which these perceptions lack, or temporal relations, which these two perceptions share. So in brief: if A and B both simultaneously feel something, according to Garrett, Hume would not be able to explain the fact that one of these feelings is A’s, the other B’s (Garrett 1981: 350-354).

This is a powerful criticism and it is not immediately clear how Hume would answer it. However, the trouble with this interpretation is that it is not clear where Garrett locates the alleged inconsistency that Hume complains of. Although Hume is genuinely dissatisfied with the ability of causation and resemblance to unite our perceptions, we are yet to find the most important reason for that dissatisfaction.

In a similar vein, David Pears (1993) claims that Hume’s position is incapable of

---

<sup>3</sup> Barry Stroud (1977) is a milestone of this interpretation. Peter Kail (2007) and Don Garrett (1981; 2011) also belong to this line of interpretation. Garrett calls this view “metaphysics of bundling”.

---

<sup>4</sup> Hume asserts that many kinds of perceptions are “no where”, just prior to the section “Of personal identity” (T253), because in his view only visual and tangible perceptions have genuine spatial locations; passions, sentiments, and even sounds, tastes, and smell do not.

explaining the peculiarities of ownership of mental objects. He points out that if Hume's view is incapable of explaining this, "his theory cannot draw the boundary between one mind and another in the right place" (Pears 1993: 295). Pears views that this created two related difficulties for Hume.

First, instead of adopting the axiom, "I feel it, it is mine," he tried to base ownership on contingent relations between perceptions; but ... his relations hold not only between the contents of a single mind but also between the contents of different minds. It is this difficulty that would be circumvented if the contents of each mind really were undetectably related to a separate nuclear self, which would provide a unique identification of the owner. The second more radical difficulty was that the three relations which, according to him, contingently connect the perceptions of a single mind, cannot possibly serve as a basis for the strong modal denial of alternative ownership; and it may be that this difficulty too would vanish if Hume could show that each person's perceptions were anchored by a separate nuclear self. But, of course, he found this idea unthinkable. (Pears 1993: 294).

The individuation of bundles is surely a problem for Hume. Stroud presents his interpretation of Hume's self-doubts as follows:

There is nothing in any perception, considered in itself, which implies the existence of any other perception, or of anything else whatsoever, and so there is nothing intrinsic to any perception that connects it with some particular series rather than another. So why do perceptions present themselves, so to speak, in discrete, separate bundles? [...] If Hume were sensing his reliance on an inexplicable 'fact' about perceptions, as I have suggested, it would be natural for him to express his quandary by saying "Did our perceptions either inhere in something sim-

ple and individual, or did the mind perceive some real connexion among them, there would be no difficulty in the case". (Stroud 1977: 138-9)

Then Stroud answers the above question in these words: "to say it is 'inexplicable' for Hume is to say that it is inconsistent with the theory of ideas which he takes to be the only way to make sense of psychological phenomena" (Stroud 1977: 140). If mental events consist in nothing more than "perception", but we cannot find intelligible connection or relation between a particular perception and a particular mind, then the theory should take perceptions as completely separated from the minds that have them, or regard them as a single large bundle.

I think that this is a trenchant criticism. And Hume may have in mind this kind of difficulty when he pronounces himself arrested in a "labyrinth". I am not convinced that they are the whole of his problem, for Hume has a number of resources available to him to account for these apparent counterexamples.

In order to solve the problem of individuation, we should take notice of the two views of personal identity that appear to be implied in Hume's remark on the distinction between "personal identity, as it regards our thought or imagination, and as it regards our passions or the concern we take in our selves" (T253). Robert S. Henderson (1990: 36-37) puts it as follows:

Hume's account of our inclination to ascribe identity to self is not simply a defense of mental-identity which is what might be suggested by reading the first book in isolation. As regard the passions, and especially pride and humility, ascriptions of self-identity depend on these evaluations which concern *the qualities of our mind and body, that is, self.*

If this reading is plausible, the above passage (T253) is a clear statement that personal identity is not simply a question of mental identity, which is “narrowly restricted” in Book 1. In order to make a distinction between the self of Book 1 and that of Book 2, the body is important to explain the self. Marina F. Spada (1998) emphasizes on the importance of the body. According to her, “the relation between the self and the body marks a difference between Book 1, and Book 2 of the *Treatise*” (1998: 196). We may suppose that “our body evidently belongs to us”; but in Book 1 (T190-1), Hume’s point is this:

’Tis not our body we perceive, when we regard our limbs and members, but certain impressions, which enter by the senses; so that the ascribing a real and corporeal existence to these impressions, or to their objects, is an act of the mind as difficult to explain, as that [of the external existence of our perceptions].

In Book 2 Hume states in a different way:

Bodily pains and pleasures are the source of many passions, both when felt and consider’d by the mind; but arise originally in the soul, or in the body, whichever you please to call it, without any preceding thought or perception. A fit of gout produces a long train of passions, as grief, hope, fear; but is not deriv’d immediately from any affection or idea. (T276)<sup>5</sup>

Here Hume seems to be talking about the *self* composed of body and mind from the perspective of common sense, with which we are well acquainted from our ordinary life<sup>6</sup>: “pride and humility have the qualities of our mind and body, that is

*self*, for their natural and more immediate causes” (T303). Hence it can be said that as John Bricke remarks, “a theory of the self that is adequate to the understanding of the emotions must [...] treat the self as a compound of mind and body” (1980: 99). If we admit that the *self* is composed of body and mind from the perspective of common sense, with which we are well acquainted from our ordinary life, the individuation problem can be solved because we can demarcate A’s and B’s perceptions by virtue of their having different bodies.

### ***1-2. The Need for a Continually Existing Self***<sup>7</sup>

Another strand of interpretation is that Hume needs “a genuinely enduring *empirical self*” supporting the associative mechanisms that explain his theory of mind. Jane L. McIntyre remarks it in these words: “the concept of a self that is *affected by experience* and therefore must *persist through experience* is precisely the concept of the self that *cannot* be accounted for in the context of the theory of ideas presented in the *Treatise*” (McIntyre 1979: 82). Why not? Both *because* that self seems to be an active agency (Kemp Smith and Robison), or a set of dispositions (Wolff and Nathanson) in addition to perception-bundles, and further *because* the “explanation of belief (in a continually existing self) which Hume is committed to seems itself to presuppose a *continually existing self*” (Passmore and MacNabb).<sup>8</sup> With variations in detail,

<sup>7</sup> Those interpreters, who belong to Group 2, according to Ellis’s classification, find Hume’s problem in what Hume calls “associationalist program”.

<sup>8</sup> See Kemp Smith (1941), also Robison (1974), Wolff (1966), Nathanson (1976), Passmore (1952), and MacNabb (1966).

<sup>5</sup> I owe this quotation to Spada (1998), pp. 196-197.

<sup>6</sup> I owe this interpretation to Capaldi (1985: 280-281).

MacNabb (1966: 151-152) and Passmore (1952: 82) offer that Hume's problem is as follows: Hume's position demands a *genuine* self rather than a merely fictitious self yet, lacking one, it is inconsistent.

According to Hume's view of causation, causality is not a relation we perceive between objects; rather we regard a pair of objects as related as cause and effect. When we have observed a constant conjunction of similar pairs of contiguous objects, our experience shows us that they have been connected in the past. But it cannot tell us that they will continue to be connected in the future. As a result of this observed constant conjunction we are disposed to expect the second member of the pair on perceiving the first. Hence, Hume concludes that the necessary connection that we attribute to causally connected objects is a projection onto them due to our mind's disposition to connect its ideas (T165-7). Now in order for us to have the belief that my perceptions, as the bundled elements which constitute an episodic self, are causally linked (Book 1), they would have to exhibit a multitude of long-standing constant conjunctions. But in order to see a multitude of constant conjunctions, there must be a self who observes that conjunction. Hence Hume's argument in Book 1 seems to be circular: it presupposes the existence of the enduring self which it aims to establish.

The circularity problem can be avoided if we consider Hume's remarks in the Appendix. Hume says: "did our perceptions inhere in something simple and individual, or did the mind perceive some real connexion among them, there wou'd be no difficulty in the case" (T636). In other words, Hume sees two ways of providing an account of a unified self, that is, as a bundle of perceptions

which is unified at one point in time, and as unified across time. One way of doing this would be if the individual perceptions were themselves the intentional objects of a "meta-self", extrinsic to the bundle of perceptions. This would be a transcendental account of the self, similar to that of Kant's when he claims that the "I think" accompanies all our judgments. The alternative is to find some intrinsic, as opposed to extrinsic, connection among the perceptions. Since Hume's metaphysics cannot allow for either possibility, he is left an insoluble problem. Hence Hume says:

I must plead the privilege of a sceptic, and confess that this difficulty is too hard for my understanding (T636).

The trouble with this interpretation is that it seems to rest on a bit of "textual magic". For these commentators think of Hume's problem as being a conflict between a Newtonian-Gravitational model and a Kantian model. However, it seems to me that Hume gives us not the slightest reason here or elsewhere to suppose that he sees this conflict.

### ***1-3. Insufficiency of the associative principles of resemblance and causation***

The third group is concerned with Hume's genetic explanation of the idea of the self, in which Ellis classifies his own interpretation (Ellis 2006: 195-232). According to this group, Hume has not troubled with the theory of ideas he uses throughout the *Treatise*. Instead, "the problem is that, even if this theory would remain unthreatened by his account of personal identity, the particular genetic explanation Hume advances

in T 1.4.6 (on the basis of this theory) is faulty” (Ellis 2006: 200). On this account, Hume’s dissatisfaction is concerned with the inadequacy of resemblance and causation to yield the idea of *my* self. According to Ellis, Hume realizes that the associative principles of resemblance and causation are inadequate to generate *the idea* of the self because he has already rejected that we have an impression of the self. According to Hume’s theory of ideas, Ellis argues, all ideas must be derived from impressions, so for Hume, a process of reflective review could never give rise to the fictitious idea of a *self* in terms of associative principles of resemblance and causation. Hence, Ellis claims that “this is at the heart of Hume’s worry in the Appendix” (Ellis 2006: 200). Actually, the problem with Ellis’s account is that Hume never thinks of the idea of the mind or self as “fictitious”. It is not the idea of the *mind* or *self* that he calls “fictitious”, only the mind’s simplicity at one time and identity through time. It is only an impression of the self as *something simple and identical* or as *distinct from our perceptions* that Hume claims not to find by introspection. Moreover, Hume never denies that there are perceptions of the self from which the idea of the self may be derived. On the contrary, he claims explicitly that “the idea, or rather impressions, of ourselves is always intimately present with us” (T317). One might wonder why Hume corrects himself by saying “the idea, or rather impressions, of ourselves is always intimately present with us”. Hume makes this claim when he discusses “sympathy” in Book 2 (T 2.3.2.11). Hume explains how the mechanism of sympathy occurs in two stages: Firstly, the observers should infer the agent’s passion by its causes or effects because it does not “discover itself

immediately to” them. The second stage is to convert the idea of the agent’s passion into the vivid impression. Here we can see that in order for us to feel sympathy with others, “force and vivacity” originated from the impression should be entered into the idea of other’s passion. Hume says as follows (T317):

When any affection is infus’d by sympathy, it is at first known by its effects, and by those external signs in the countenance and conversation, which convey an idea of it. This idea is presently converted into an impression, and acquires such a degree of force and vivacity, as to become the very passion itself, and produce an equal emotion, as any original affection.

In light of this, Garrett argues that “Hume is saying only that the mental mechanism of sympathy is triggered by the recognition of resemblance between ourselves and others, and that the force and vivacity needed for this mechanism are derived from our impressions or memories of ourselves – in a sense in which there are such impressions and memories” (1981: 343). In this respect, Norman Kemp Smith (1941) remarks that it is “an awareness of personal identity” that Hume recognizes to be inconsistent with the two principles (Kemp Smith 1941: 555-558). Kemp Smith thinks that there is the inconsistency *between* Hume’s denial in Book 1 of the *Treatise* that we have an impression of the “self” and Hume’s appeal to the impression, in Book 2, “the idea, or rather impressions, of ourselves is always intimately present with us” (T317). Given this, according to Kemp Smith, Hume’s difficulty is that Book 2 requires an awareness of personal identity that his own theory in Book 1 will not allow and his second thoughts in the

Appendix are concerned with his realization of this. I doubt that this is the reason for Hume's dissatisfaction, because Hume has resources available to him to escape from this objection.

According to Hume, distinct perceptions are distinct existences. The thoughts, sensations, and emotions which constitute the self exist independently of reflection. So, if Hume admits "reflection" on mental contents as synthesizing the activity of the mind, it can "produce a sense of a unified and continuing self by producing the feeling that these conjunctions are necessary and that the conjoined items are connected" (Swain 2006: 141). One reason to reject this reading is that, as Garrett points out, Hume thinks that "personal identity 'arises from consciousness', which is itself 'nothing but reflected thought or perception' – that is, a matter of memory" (Garrett 1981: 340). According to Garrett, Kemp Smith does not recognize that "reflected thought or perception" refers to *memory*, and claims that the nature of this "reflection" is the subject of Hume's despair (Garrett 1981: 340).

Having understood the third group that Ellis classifies in above, we can see that Ellis does not think that Hume's concern is with the metaphysical question of how perceptions are actually related so as to constitute bundles. Hence, he interprets Hume's worry in the Appendix as referring only to the operations of the psychological principle of association, and not to metaphysical principles of unification. Given this, Hume's despair in the Appendix, according to Ellis, is derived from his inability to explain how the psychological principles of association generate the ascription of identity to the self and does not concern any metaphysical principle of unification. If this were right,

as Garrett points out, Hume would have been concerned only with the origin of the idea of the self, and he could have called the section "Of the idea of self", just as he calls the other sections of the Treatise: 1.2 "Of the ideas of space and time", 1.3.2 "Of probability; and of the idea of cause and effect", and 1.3.14 "Of the idea of necessary connection" (Garrett 2011: 27).

If the self is a bundle, and its parts "form a whole only by being connected together" (T635), we can say that a satisfactory account of the self should explain what actually binds together the perceptions that constitute a self. Since Hume argues that it is impossible to explain the connection, provisionally I suggest his disillusionment might be that he cannot explain what the connection or tie really is among the perceptions that are felt to be parts of this identity.

Having established Hume's difficulty concerning personal identity in the Appendix in this way, what then is the inconsistency that Hume worries about in the Appendix? According to Waxman (1992), Hume's inconsistency lies between Hume's conclusion in T 1.4.6 concerning the impossibility of any metaphysical explanation of the connecting principle for the self's distinct perceptions and Hume's earlier claims in T 1.4.5 that the "intellectual world, tho' involved in infinite obscurities, is not perplex'd with any such contradictions, as those we have discover'd in the natural" (Waxman 1992: 223).<sup>9</sup> When Hume reviews this on the section of personal identity in the Appendix, he realizes that (T631):

I had entertain'd some hopes, that however deficient our theory of the intellectual

---

<sup>9</sup> Corliss Swain also finds Hume's inconsistency in this way (see Swain 2006: 133-50).

world might be, it would be free from those contradictions, and absurdities, which seem to attend every explication, that human reason can give of the material world. But upon a more strict review of the section concerning personal identity, I find myself involv'd in such a labyrinth, that, I must confess, I neither know how to correct my former opinions, nor how to render them consistent.

In this passage, Hume seems to say that his theory of personal identity involves unavoidable inconsistencies. This inconsistency derives from his earlier claim that accounts of the intellectual world (or the self) are special in that they alone are, or can be, free of contradiction. When he reviews on the section of the personal identity, he finds it defective.

***1-4. The inconsistency between Hume's conclusion in T.1.4.6 concerning "metaphysics of bundling" and his earlier claim in T.1.4.5 concerning "intellectual world"***<sup>10</sup>

In a similar vein, Corliss Swain (2006) finds the source of Hume's inconsistency in Hume's claim that "the self is a system of connected perceptions but goes on to say that the only knowable link between the supposedly connected parts is a link between *ideas* of these parts rather than a link between the parts themselves" (Swain 2006: 140). We can find the textual evidence of this interpretation in the Appendix when Hume (T635) summarizes his discussion of Book 1:

---

<sup>10</sup> Ellis attributes this interpretation to Kemp Smith (1941), Penelhum (1951), and Swain (2006). Ellis claims that this interpretation generally gives no explanation of how Hume's problem arises specifically when he comes to explain the 'principles, that unite our successive perceptions in our thought or consciousness.'

If perceptions are distinct existences, they form a whole only by being connected together. But no connections among distinct existences are ever discoverable by human understanding. We only feel a connection, or a determination of the thought, to pass from one object to another.

Since this feeling is just one perception among the many that make up the self, it cannot connect distinct perceptions. The feeling itself arises from the regularities we find in the psychological realm.

In the Appendix, Hume claims that the attempt to explain "the principles, that unite our successive perceptions in our thought or consciousness" is hopeless. He also says that when he tries to explain it, he finds that his account is "very defective" (T635). Is this a metaphysical agony? According to Hume, all the metaphysical explanations require that the mind conceives of a connection between distinct existences. However, if the mind never conceives of a connection between distinct existences, then no satisfactory metaphysical account of any composite thing is possible. When Hume realizes in the Appendix that the same "contradictions" can be found in philosophical theories of the intellectual world as are found in the natural world, this means that skepticism about metaphysical theories of both the mind and the external world is even stronger than he had thought before (in Book 1). This skepticism derives from the principle that the mind never perceives connections among distinct existences.

## **2. Alternative Interpretation**

According to Daniel Dennett, "Hume wisely shunned the notion of an inner self that would intelligently manipulate the

ideas and impressions”. This left him only one alternative, if he wished to avoid a mysterious duplication of personal-level properties. He was left, Dennett suggests, “with the necessity of getting the ideas to ‘think for themselves’”. Even though “his associationistic coupling of ideas and impressions, his pseudo-chemical bonding of each idea to its predecessor and successor, is a notorious non-solution to the problem” (Dennett 1978: 102). Dennett thinks that Hume had no alternative but to take this seriously. What, then, is Dennett’s alternative? According to this view, the alternative seems to be to posit a sub-personal interpreter, a homunculus, as the subject of the representation. However, it seems that to explain the intelligence of people by positing intelligent homunculi is to embark on a philosophically pointless regress. As John Biro points out, “unless we are ultimately able to get rid of such a homunculus by explaining how its functions can be carried out by dumb physical components of the system, we are left with an ‘exempt agent’ whose intelligence is unexplained” (1992: 54). Faced with this criticism, Dennett says that the homunculus model does *not* embark on circularity or infinite regress. For:

Homunculi are *bogeymen* only if they duplicate *entire* the talents they are rung into explain... if one can get a team or committee of *relatively* ignorant, narrow-minded, blind homunculi to produce the intelligent behavior of the whole, this is progress. (Dennett 1978: 123)

Now in what sense does Dennett think that it can be progress to invoke homunculi who are “more stupid” than the person or

a homunculus whose attainment needs explanation? He answers as follows:

The AI programmer begins with an intentionally characterized problem, and thus frankly views the computer anthropomorphically; if he solves the problem he will say that he has designed a computer that can [for instance] understand questions in English. His first and highest level of design breaks the computer down into sub system, each of which is given intentionally characterized tasks; he composes a flowchart of evaluators, remembers, discriminators, overseers and the like. These are homunculi with a vengeance; the highest level of design breaks the computer down into a committee or smaller homunculi, but more important, into *less clever* homunculi. When the level is reached where the homunculi are no more than adders or sub tractors, by the time they need only the intelligence to pick the larger of two numbers when directed to, they have been reduced to functionaries “who can be replaced by a machine”. The aid to comprehension of anthropomorphising the elements just about lapses at this point, and a mechanistic view of the proceedings become workable and comprehensible. (Dennett 1978: 80)

Although Dennett’s suggestion improves upon Hume’s explanation of the belief in the self, it seems to me that Dennett’s argument can scarcely be used to support Hume’s view. This is not achievable, unless the terminology of AI can be rigorously and legitimately used and adapted to Hume’s view on selves and their various sub-personae in substantially the same sense as it used for computers and their homuncular sub-systems. Nonetheless, I think, that the attribution of a person’s intelligence to a set of sub-persons (or sub-persona processes) is a possible model of explanation that can eliminate Hume’s difficulty.

## Conclusion

So far we have examined a number of views on why Hume feels trapped in labyrinth in the Appendix. However, no consensus has emerged. It seems to me that we cannot give an exact answer to the issues raised above, for Hume never says explicitly what is bothering him. However, we could suggest what Hume *ought to say* about these questions. I have suggested a proposal that at least can meet these demands by considering Daniel Dennett's (1978) alternative perspective on Hume's Problem.

In short, we can say that Hume's despair in the Appendix is concerned with finding the ultimate principles of connection among our distinct perceptions that constitute the self. In order to avoid this difficulty, Hume correctly sees inherence and real necessary connections as the only apparent ways out. But, from Hume's empiricist's perspective, it is impossible to find those connections veiled behind impenetrable obscurity.

The upshot of this is that causal explanation (implying necessary connection) is no longer possible when Hume cannot find the ultimate sources and the principles that connect the perceptions that make up the self. This is because cause and effect are not directly connected, but are connected

in the minds of perceivers. Thus, the causal link is a link in perceivers' minds between their perceptions of the causally related objects. Although Hume is a skeptic about a metaphysical account of causation in denying a necessary connection between causally related objects, such skepticism is compatible with belief in causation and an indirect connection that exists only in the minds of perceivers when they observe the constant conjunctions. Hume's skepticism about a metaphysical account of connecting principles does not undermine our belief in the connection. Our beliefs in connections among perceptions that constitute the self, according to Hume, are inevitable and indispensable in the conduct of life. They can be justified in the absence of a satisfactory metaphysical account. However, in order for the perceiver to believe in causation and personal identity, there must be a self who observes those constant conjunctions and feels those successive perceptions. In rejecting the metaphysical substantial self and any sort of real connections among perceptions, he is left bereft of any explanation for the nature of personal identity and causation. Hence, we can say that his problem arises from his view on the realm of the mental, which empiricism alone cannot fully explain.

## REFERENCES

- Ainslie, D., 2001. Hume's reflections on the Identity and Simplicity of Mind. *Philosophy and Phenomenological Research* 62 (3): 557–78.
- Baier, A., 1991. *A Progress of Sentiments: Reflections on Hume's Treatise*. Cambridge, Mass.: Harvard University Press.
- Bricke, J., 1980. *Hume's Philosophy of Mind*. Edinburgh: Edinburgh University Press.
- Biro, J., 1992. Hume's New Science of the Mind. In: D. F. Norton (ed.), *The Cambridge Companion to Hume*, Cambridge: Cambridge University Press, pp. 33–63.
- Capaldi, N., 1985. The historical and philosophical significance of Hume's theory of the self. In: A. J. Holland (ed.), *Philosophy, Its History and Historiography*, Dordrecht: Reidel, pp. 271–285.
- Dennett, D., 1978. Artificial Intelligence as Philosophy and as Psychology. In: *Brainstorm: Philosophical Essays on Mind and Psychology*, Cambridge, Mass.: MIT Press.

Ellis, J., 2006. The Contents of Hume's Appendix and the Source of His Despair. *Hume Studies* 32.2: 195–233.

Garrett, D., 1981. Hume's Self-Doubts about Personal Identity. *The Philosophical Review* 90 (3): 337–358.

Garrett, D., 2011. Rethinking Hume's Second thoughts about Personal Identity. In: J. Bridges, N. Kolodny, and W. Wong (eds.), *The Possibility of Philosophical Understanding: Essays for Barry Stroud*, New York: Oxford University Press, pp. 15–42.

Hume, D., 1978. *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2<sup>nd</sup> ed. Revised by P. H. Nidditch, Oxford: Clarendon Press.

Henderson, R. S., 1990. David Hume on Personal Identity and the Indirect Passions, *Hume Studies* 17: 33–44.

Kail, P. J. E., 2007. *Projection and Realism in Hume's Philosophy*, Oxford: Oxford University Press.

Kemp Smith, N., 1941. *The Philosophy of David Hume*, London: Macmillan.

McIntyre, J. L., 1979. Is Hume's Self-Consistent? In D. Fate Norton, N. Capaldi, and W. L. Robison (eds.), *McGill Hume Studies*, San Diego, CA: Austin, pp. 79–88.

MacNabb, D. G. C., 1966. *David Hume: His Theory of Knowledge and Morality*. Oxford: Blackwell.

Nathanson, S., 1976. Hume's Second thoughts on the Self. *Hume Studies* 2: 36–45.

Noxon, J., 1973. *Hume's Philosophical Development*. Oxford: Clarendon Press.

Passmore, J. A., 1952. *Hume's Intentions*. Cambridge: Cambridge University Press.

Pears, D., 1993. Hume on Personal Identity. *Hume Studies* 19: 289–300.

Robison, W. L., 1974. Hume on Personal Identity. *Journal of the History of Philosophy* 12: 181–93.

Robison, W. L., 1979. In Defense of Hume's Appendix. In D. Fate Norton, N. Capaldi, and W. L. Robison (eds.), *McGill Hume Studies*, San Diego, CA: Austin, pp. 89–99.

Roth, A., 2000. What was Hume's Problem with Personal Identity. *Philosophy and Phenomenological Research* 61(1): 91–114.

Spada, M. F., 1998. *Space and the Self in Hume's Treatise*. Cambridge: Cambridge University Press.

Stroud, B., 1977. *Hume*. London: Routledge & Kegan Paul.

Yang, S., 2006. *Emotion and Intentionality in Understanding Values and Personal Identity: A Philosophical Study of Emotion from a Humean Point of View*, PhD Thesis, University of Durham.

Swain, C. G., 1991. Being Sure of Oneself: Hume on Personal Identity. *Hume Studies* 17: 107–124.

Swain, C. G., 2006. Personal Identity and the Skeptical System of Philosophy. In: S. Traiger (ed.), *The Blackwell Guide to Hume's Treatise*, Malden, Mass.: Blackwell.

Waxman, W., 1992. Hume's Quandary Concerning Personal Identity, *Hume Studies* 18 (2): 233–53.

## HUME'O ABEJONĖS DĖL ASMENS TAPATUMO

### Sunny Yang

**Santrauka.** Šiame straipsnyje svarstoma, kaip Davidas Hume'as galėtų išvengti savo intelektinių kančių dėl asmens tapatumo *Traktato* Priede. Visų pirma, straipsnyje siekiama suprasti Hume'o nusivylimo, bandant *Traktato* Priede paaiškinti tai, kas jam iš tikrųjų rūpėjo, šaltinis, ir pasiūlomos kelios interpretacinės linijos. Neseniai Jonathanas Ellisas išskyrė keletą galimų Hume'o keblumų supratimo būdų. Šiame straipsnyje iš keturių Elliso išskirtų paaiškinimų grupių detalai aptariami trys jo menkai eksplicuoti dalykai ir sprendžiami kai kurie svarbūs Elliso praleisti klausimai. Galiausiai pasiūlomas alternatyvus Hume'o sunkumo perskaitymas, grindžiamas Danielio Dennetto idėjomis šiuo klausimu, ir pateikiamas pasiūlymas dėl to, ką Hume'as būtų galėjęs pasakyti apie šias problemas.

**Pagrindiniai žodžiai:** asmens tapatumas, Hume'o abejonės, jungiamieji principai, Hume'o nenuoseklumas, homunkulų modelis

*Įteikta 2018 05 31*

*Priimta 2018 08 13*