



Hume's Labyrinth Concerning the Idea of Personal Identity

Donald L. M. Baxter

Hume Studies Volume XXIV, Number 2 (November, 1998) 203-234.

Your use of the HUME STUDIES archive indicates your acceptance of HUME STUDIES' Terms and Conditions of Use, available at

<http://www.humesociety.org/hs/about/terms.html>.

HUME STUDIES' Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the HUME STUDIES archive only for your personal, non-commercial use.

Each copy of any part of a HUME STUDIES transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

For more information on HUME STUDIES contact humestudies-info@humesociety.org

<http://www.humesociety.org/hs/>

Hume's Labyrinth Concerning the Idea of Personal Identity

DONALD L. M. BAXTER

I. Introduction.

In *A Treatise of Human Nature* Hume argues that the self is really many related perceptions, which we represent to ourselves as being one and the same thing. In a perplexing Appendix he repudiates this account. Scholars have found various reasons for repudiation, but none that are certifiably Hume's. I propose that his reason is an inconsistency between his skeptical arguments that the self is really many perceptions and his naturalist explanation of how we come falsely to attribute identity and simplicity to these many perceptions—how we come to have an idea of a simple, identical self. The inconsistency threatens the delicate alliance between Hume as skeptic and Hume as mental naturalist.

As I read the Appendix, the main points are these: the self is many distinct perceptions. We think of these distinct perceptions as one and the same thing. Hume had hoped this thought could be explained by the fact that we merely *feel* a connection between the perceptions when reflecting on them. But he discovers that this won't work and he can think of only two possible explanations that will: either the perceptions have a real connection that can be copied in the ideas reflecting them, or only the reflecting ideas themselves have a real connection. But this real connection is identity. So either our many perceptions are identical as are the ideas reflecting them, or only the many ideas reflecting them are identical. Either option is inconsistent.

The source of this inconsistency is left implicit. I suggest that the source is an often-implicit part of Hume's theory of representation. Concerning perceptions and their objects he holds that, roughly, representation requires resemblance. It will turn out that only a many can represent a many and only a one can represent something *as* one. So for the many distinct perceptions to be represented *as* one and the same, the many distinct ideas that reflect them must *be* one and the same. But the distinct cannot be identical, so Hume cannot both posit a self that is many and an idea of the self as one.

Even worse, his account of such an idea turns out to be an affirmation of the unitariness of the self. The idea itself would be the self. So Hume the skeptic says the self is many and Hume the naturalist says it isn't.

I have said that in the Appendix Hume finds himself committed to the following: the ideas reflecting some successive past perceptions have to be connected into a simple identical self in order to represent the past perceptions as connected into a simple identical self. Whether this is really what Hume thought, I don't know. That it is the only possible good reading, I don't pretend. But of the extant interpretations mine is the one most faithful to the text of Hume's Appendix and the one most careful to explain why Hume used the words he did, or so I will argue.¹

The main trouble with most other interpretations is that they focus on the question "What is the problem Hume found in his bundle theory of the self?" The various answers, though illuminating, stray from the specific concern of the Appendix, which is better revealed by asking "What is the problem Hume found in his bundle theory of *the idea* of the self?"²

The focus on the former question is understandable. At *Treatise* 635,³ when announcing the problem, Hume draws attention to the "principle of connexion" that binds the "loosen'd" perceptions together. It is natural to think the locus of concern is how some relation between the perceptions, or their relation to something else, bundles the perceptions into a self. This is reinforced by the emphasis in I iv 6, "Of personal identity." There he explains how, when reflecting upon the perceptions, their relations of resemblance and causation cause an easy transition of the ideas of them in the imagination, thus causing the imagination to attribute simplicity and identity to them—that is, to take them to be numerically identical with each other. These relations between the perceptions and their relation to the imagination seem to be Hume's main concern as he proceeds, in his words, "to explain the nature of personal identity." That identity, he insists, "is only a fictitious one" (T 259). Instead of a single self there are really the many related perceptions bundled by the imagination. However Hume's focus in the Appendix is different. There he is not so much concerned with these relations between the perceptions. He is more concerned with the "connexion" that cannot be "discovered" but is only "felt" between them

when reflecting on them. This connection is identity, as he explicitly says at T 259. But since it cannot be discovered, the emphasis shifts to its being felt. Now a connection felt on reflection between things—even between perceptions—is a connection felt during the reflection, between the *ideas of* those things (T 167–169). Normally such a felt connection is just the “easy transition” of these ideas. In the Appendix, however, there is a problem with “feeling” identity as a basis for attributing identity. It is a problem that the standard appeal to an easy transition cannot solve. Note that those ideas had while reflecting on the perceptions composing the self are what compose the idea of the self. So Hume's concern in the Appendix is with the connection within the idea of the self.

If I am right, then interpretations of the Appendix concerned with *relations* bundling perceptions into selves are off the mark. These include those of Basson (131–133), Beauchamp, Flage (149–150), Garrett, Grice and Haugeland, Loeb, McNabb (149–151), Patten, Pears, Robison (“Defense”), and Stroud (138–139).⁴ Likewise, interpretations concerned with a *non-Humean* selves bundling perceptions into selves are off the mark. These include the interpretations of Johnson (297–298), Kemp Smith (73), Passmore (82–83), Price (5–6), and Robison (“Personal Identity”).⁵ All of these interpretations, even those focused on problems with getting the idea of the self, miss Hume's main concern—the connection within the idea of the self.⁶

Interpretations in the first group are right insofar as they take seriously Hume's insistence that any self is just many perceptions. Those in the second group give this up too soon. They hold that only a unitary, active self not composed of perceptions (i.e., a non-Humean self) could smoothly transit the many associated perceptions and confound them with a single thing. But this assumption just begs the question against Hume's view that the self is many passive perceptions. As several commentators have pointed out, for Hume association, transit, and confounding are just a matter of which perceptions tend to accompany or follow one another; no further explanation is possible or necessary.⁷

Despite insufficient argument, the interpretations in the second group are right insofar as they have Hume questioning his insistence that the self is many. He does indeed worry that his account of attributing unity commits him to a unitary self, but again this is because he worries about the connection within the idea of the self. It is not that he thinks there must be something unitary, not composed of perceptions. It is rather that the idea of the self, with a real connection between component ideas, would be just the sort of thing Hume took the self as traditionally conceived to be—many perceptions that are somehow also identical.

The problem is with the fictitious attribution of identity to the many perceptions. There is a third group of interpretations, most not specifically of the Appendix, that can be used to argue that the identity attributed is not

fictitious, so there is no false attribution to explain.⁸ This group includes Ashley and Stack, Biro, Kemp Smith (96-98, 499-502), Laird (171-172), MacNabb (147-148, 151), Penelhum ("Personal Identity" and "Revisited"), Perry (26-30), and Swain.⁹ This group roughly divides into those who think identity, not perfect but imperfect, is rightly attributed either to the distinct perceptions or to the bundle of them as a whole, and those who think identity *simpliciter* is rightly attributed to the bundle of successive perceptions as a whole.

The first sub-group thinks that perfect identity is the sort of identity characteristic of unchanging steadfast objects, while imperfect identity is the sort characteristic of successions. But this is either confused or the position of the second sub-group. If imperfect identity is a relation between distinct things, it is not identity no matter what you call it.¹⁰ If it is a relation a succession has to itself then it is just identity regardless of that to which it is being applied. In neither case are there two kinds of identity, perfect and imperfect. Nor did Hume believe in two kinds of identity. For him, perfect identity is just identity. Imperfect identity is just identity naturally but falsely believed to hold between distinct objects because of the relations between them.¹¹ It is, so to speak, imaginary identity found in "instances of this tendency of relation to make us ascribe an *identity* to *different* objects..." (T 204).

In getting the idea of the self, what are the "different objects" to which identity is falsely ascribed? They are the "successive perceptions"¹² (T 253). The second sub-group neglects the point that the relevant relata of the attributed identity are perceptions, not bundles of them.¹³ Perhaps Hume could have given identity conditions for bundles of perceptions, *qua* bundles. But his concern is to explain why, when introspecting many successive things, we take there to be a single steadfast thing *which is not a bundle at all*.¹⁴ Hume thinks it is because we take distinct perceptions to be identical (while half recognizing the absurdity of this) that we come up with the idea of a single thing to which they belong. That we have such an idea, Hume takes for granted. So the false attribution of identity must be explained. Hume finds that this falsely ascribed identity—i.e., imperfect identity—requires an idea of the self that is many ideas somehow perfectly identical with each other.

II. Denying the Identity and Simplicity of the Self.

The relevant part of the Appendix comprises twelve paragraphs, starting with "I HAD entertain'd some hopes...." and ending with "those contradictions" (T 633-636).

Hume begins the first paragraph by giving up his conviction that our ideas of "the intellectual world" would be free of the inconsistencies in our

ideas of "the material world." He expresses this conviction at the beginning of "Of the immateriality of the soul" (T 232; see also 366). I speculate that the root of this conviction is what I call the Cartesian Assumption that the mind and its modes are as they appear to careful introspection. In the case of bodies he is cautious about drawing conclusions about objects based on our knowledge of our impressions (T 241–242). Why then, from the fact that when he reflects on himself he perceives perceptions and only perceptions, does Hume unhesitatingly "venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions..."?¹⁵ (T 252). He seems to be assuming that, while external objects may ultimately be mysterious to us, the mind or self is what it is perceived to be.

Hume next confesses that he cannot render his former opinions *consistent*. He has found a contradiction. The former opinions are those presented in I iv 6: the opinion that the self is really many distinct things, and the account of how we come falsely to believe it to be simple and identical. He resolves to "propose the arguments on both sides." First he gives "those that induc'd me to *deny* the strict and proper identity and simplicity of a self or thinking being" (T 633, my emphasis). I suggest that next he gives the argument that leads him to *affirm* the identity and simplicity of the self. It comes in the course of explaining "the principle of connexion, which binds them [the loosen'd perceptions] together, and makes us attribute to them a real simplicity and identity" (T 635). Thus it is the argument *for* his explanation of the attribution that raises the consideration *against* the denial of identity.

A note on usage: Hume uses 'identity' the way we would use 'identity through time'. He says that identity is "common to every being, whose existence has any duration" (T 14). Were he using our current unrestricted concept of identity he would have said that identity is common to every being, whether or not it has duration. Elsewhere Hume claims that the idea of duration is integral to the idea of identity¹⁶ (T 200–201).

Denying something's identity, for Hume, is affirming that "it" is distinct things in succession, rather than one continued thing. Through time, "it" is many rather than one. This is why Hume makes clear he is denying "the strict and proper identity" (T 633). This "strict" identity is in contrast to "imperfect identity," in which case we naturally but falsely attribute strict identity to related objects in succession (T 256). That is, we regard many successive things as identical with each other, and so as one steadfast thing. Hume's denial of the strict identity of the self is part of an affirmation of its imperfect identity, i.e., false but supposed identity.

Likewise denying something's simplicity is affirming that it is a complex thing, which for Hume is to affirm that "it" is many co-existent things, rather than one single thing. He equates simplicity with being one and contrasts it with composition. Speaking of "the several distinct sensible

qualities, of which objects are compos'd" he says

But however these qualities may in themselves be entirely distinct, 'tis certain we commonly regard the compound, which they form, as ONE thing, and as continuing the SAME under very considerable alterations. The acknowledg'd composition is evidently contrary to this suppos'd *simplicity* and the variation to the *identity*. (T 219)

Regarding the object as one thing is supposing it to be simple. Its really being complex entails that "it" is really many.¹⁷ As for such an object, so for the self.

The scope of "strict and proper" quoted above, which certainly includes "identity" may well also include "simplicity"¹⁸ (T 633). Although Hume does not speak of "imperfect simplicity" he could well be thinking of it.¹⁹ Analogous to imperfect identity, imperfect simplicity could characterize cases in which many co-existent closely related things are regarded as one simple thing. At T 221 he talks about "a peach or melon" being just such a case (see also T 263). So denying strict and proper simplicity to the self is affirming its imperfect simplicity, which is false, but supposed, simplicity.

Thus in denying the strict and proper identity and simplicity of the self, Hume is affirming that "it" is many things at any given time, and many successive things through time. That is to say, "it" is not one single thing.

First are the arguments on the side of denying the self's simplicity and identity. These occur in paragraphs 2–10. The arguments are couched in terms of a denial that the self is a substance in which perceptions inhere as accidents.²⁰ In the second paragraph, he says that the terms 'self' and 'substance' must each stand for some idea. If not, then they are "altogether unintelligible." This consequence seems too strong. Berkeley had stressed in his Introduction to the *Principles* that words can have other functions besides standing for ideas.²¹ But it does not matter: Hume's main concern is not with requirements for intelligibility, but with what the ideas in fact annexed to 'self' and 'substance' are like.

First he says what they are *not* like: "we have no impression of self or substance, as something simple and individual. We have, therefore, no idea of them in that sense." I take it that he means here that we have no idea of anything which is *in fact* a simple and individual self. He certainly thinks that in *some* sense we have a fictitious idea of the self as simple and individual, for explaining the genesis of that idea is a large part of *Treatise*.²² But in the Appendix he is concerned to show that what the idea of the self is in fact *of*, is a bundle of perceptions.

In paragraphs three through five, he argues that there is no absurdity in thinking of the self as many perceptions which do not inhere in anything. Like ordinary objects, perceptions can "exist distinct and independent,

without any common *simple* substance or subject of inhesion" (T 634). The basis of his claim is an principle—taken for granted without argument—concerning identity and distinctness that I will call the Separability Principle: If things are numerically distinct then they can exist separately²³ (T 634). The argument is that our perceptions are all distinct from each other, each can exist without the others, and "without any common *simple* substance or subject of inhesion." Hume does not argue for this last directly from the Separability Principle, but he could have. Either all our distinct perceptions are identical to the substance they inhere in or not. If all are identical, then distinct perceptions are identical with the same thing, but that is absurd. So some are distinct from the substance. If distinct, then they can exist without it. So there is no common simple substance. Instead of arguing this way, Hume takes another tack. He says we all take for granted that distinct ordinary objects, such as a table or a chimney, can exist on their own. And he says that what is true of objects is also true of perceptions (T 241–242).

In paragraphs six and seven he appeals to there being no evidence of anything but perceptions when one attends to the self. When Hume reflects on himself he finds only perceptions. Their existence (based on the previous results) implies the existence of nothing else. So they are all that he can know exists when he is examining the self. So they *are* the self. "'Tis the composition of these, therefore, which forms the self" (T 634). This conclusion apparently follows only given the Cartesian Assumption that there is no more to the self than what is evident to careful reflection. Even a simple self—e.g., the single perception of an oyster—would betray no substance, no principle of simplicity. So much less would a complex self as Hume has discovered he is.

In paragraphs eight through ten he gives two more arguments and then concludes the evidence for his "former opinions." First, one's self is inseparable from one's various perceptions taken collectively. This is proven given the question-begging assumption that death is the destruction of the self and yet is nothing but the extinction of its perceptions. But the inseparable are identical, given the contrapositive of the Separability Principle. Second, the self and substance are both merely principles of unity and yet some philosophers (e.g., Locke²⁴) distinguish them. Hume again can make no sense of this and can make sense of the self only as myriad perceptions. Philosophers aided by Locke's pessimistic remarks have made it plausible that we have no idea of material substance except just several qualities in concert.²⁵ Hume is arguing that likewise we have no idea of mental substance except just several perceptions in concert. He gives the principle "*that we have no notion of it [mind], distinct from the particular perceptions*" (T 635).

"So far I seem to be attended with sufficient evidence" (T 635). These arguments satisfy Hume. They are those "that induc'd me to *deny* the strict and proper identity and simplicity of a self or thinking being" (T 633, my emphasis). But now he will give an argument *affirming* these. For he said he would "propose the arguments on *both sides*" (T 633, my emphasis). The argument comes in the course of giving the remaining part of his "former opinions"—the explanation of how we come falsely to attribute identity and simplicity to the self. Here is where he finds his account to be "very defective."

III. Affirming the Identity and Simplicity of the Self

Hume begins paragraph eleven by describing his previous arguments as having "loosen'd all our particular perceptions." In other words, he has shown that there is nothing more to the mind than these "distinct existences." Now he pessimistically proceeds to explain "the principle of connexion, which binds them together, and makes us attribute to them a real simplicity and identity" (T 635). He is assuming that there are two possible causes for our attributing simplicity and identity: either we perceive, discover, the connection between the perceptions, or else we feel it. This assumption presumably relies on his discussion of causal necessary connection. There the alternative to discovering that sort of connection is feeling it. And, of course, feeling it is not discovery at all but is rather a "spreading" on their objects of a feeling occasioned by the transition of ideas (T 167–169). This same theme appears in the discussion of personal identity, the only difference being the sort of connection (T 259–260, T 635). Yet in this new context a problem arises.

The problem begins with the assumption that "no connexions among distinct existences are ever discoverable by human understanding." In the next paragraph, the twelfth, he says if he could give this up he wouldn't be faced with the prospect of perceptions inhering "in something simple and individual." But he can't give it up; it is one of his most fundamental principles.²⁶ So he proceeds: "We only *feel* a connexion or determination of the thought, to pass from one object to another."

What is the "connexion" that is not discoverable but only feels as if it is there? He has told us at T 259: it is identity.

But, as, notwithstanding this distinction and separability, we suppose the whole train of perceptions to be united by identity, a question naturally arises concerning this relation of identity; whether it be something that really binds our several perceptions together, or only associates their ideas in the imagination. That is, in other words, whether in pronouncing concerning the identity of

a person, we observe some real bond among his perceptions, or only feel one among the ideas we form of them.

In answer to the question he says:

identity is nothing really belonging to these different perceptions, and uniting them together; but is merely a quality, which we attribute to them, because of the union of their ideas in the imagination, when we reflect upon them. (T 260)

So we attribute identity to the many successive perceptions not because they really are identical—that we can't discover—but because they feel identical. The same goes for simplicity (T 263). We attribute simplicity to the many co-existent perceptions—that is we take them to be one partless thing—not because they are something simple, but because they feel like something simple. Attributing simplicity is just like attributing identity in Hume's sense: it is attributing identity in the current sense to the many parts of a composite. That is what it is to take many coexistent things to be something partless. Thus, attributing identity and simplicity in Hume's senses to many perceptions, some co-existent, some successive, is simply to attribute to them identity in the current sense. It is to take all the distinct things, whether at a time or through time, to be the same thing. Identity, in the current sense, is the undiscovered felt connection whether Hume speaks of identity or simplicity. In any event, the problem is with explaining how the process of perceptions feeling identical and simple can serve as a basis for our attributing identity and simplicity to them. A defect in this explanation is what leads Hume to worry that he would have to have them inhere in a simple identical self.

If the connection is identity and simplicity, the "*principle of connexion*" (my emphasis) is the *cause* for the attribution of this connection. In a footnote Hume refers to T 260 where he gives the cause(s) as the resemblances and causal connections between the successive perceptions. These cause the attribution of identity by producing "an easy transition of ideas." It is this account that he finds "very defective." The problem is how the easy transition can explain the attribution.

What happens, on his account, is that one reflects on a train of past perceptions (T 635). That is, one has a current train of reflecting ideas—ideas which reflect, i.e., copy, and so are of those perceptions.²⁷ In this case these reflecting ideas are memories. He says that memory "repeats" impressions though with less vivacity (T 8). That is to say that some memories are copies of the impressions remembered. Likewise other memories are copies of ideas remembered, if we can assume memory is the faculty that produces "secondary ideas" (T 6–7). These memories, reflecting ideas, in the train are such that each feels as if it attracts the next to mind; they "naturally

introduce each other." This feeling is "spread" onto the past perceptions (T 167). Thus the past perceptions are *felt* to compose a unity because of the felt attractions between the ideas reflecting them. This felt connection between the perceptions "makes us attribute to them a real simplicity and identity" (T 635).

In the Appendix he briefly summarizes the account at T 253–254 and 259–260 of the origin of the idea of a simple, identical self.²⁸ There he says the mind feels an easy transition along a succession of related objects upon which it reflects.²⁹ As a result it takes the objects in succession to be rather a single invariable and uninterrupted thing at successive times.³⁰ More tricks of the imagination follow in Hume's account, but this much is what I want to focus on. Hume reminds us of this account in the Appendix, but in the Appendix there is something new: for the first time Hume tries to eliminate the conceit of the mind easily transiting the successive ideas which are of the train of past perceptions.³¹ He tries to say more literally what is going on, following up on things he said at T 260. His appeal to the mind there seems like an appeal to a core perceiver, unitary and active, that reflects on the train of perceptions by having a succession of ideas while having certain feelings and making certain confusions. It is much like the situation were we to "see clearly into the breast of another, and observe that succession of perceptions, which constitutes his mind or thinking principle..." (T 260). There is a clear distinction between the observer, seemingly unitary and active with ideas of the perceptions, and the bundle of perceptions observed.³¹

The question naturally arises, what constitutes the mind of the core perceiver? In the Appendix Hume tells us. The core perceiver is just the ideas it is having; there is nothing else to it. What collectively play the role of the mind of the core perceiver are just ideas in "consciousness." Consciousness is "nothing but a reflected thought or perception" (T 635). In other words, the consciousness of a perception, at least in this context, is an idea that is a copy or reflection of a previous perception. The easy transition in the core perceiver between past perceptions is really just the fact that the ideas of them in consciousness are "felt to be connected together, and naturally introduce each other." In other words, these reflecting ideas are such that the earlier ones attract the later ones to mind, accompanied by impressions of smoothness and ease.³³ These events in consciousness cause an idea that represents the past train of perceptions as being identical. To summarize, Hume thinks that a felt connection in consciousness—i.e., that between reflecting ideas—must be appealed to in order to explain the felt identity between the past perceptions.

While Hume allows that this is a surprising account, he asserts that it is consistent with the prevailing view that "personal identity arises from consciousness." Here I speculate that Hume is lumping together the views of

those who think consciousness and thought are the best *evidence* of personal identity with those who think they can *define* personal identity in terms of consciousness.³⁴ Hume likely thinks his view that consciousness causes the false belief in identity explains the appeal of these other views. Thus far his view has "a promising aspect."

But suddenly Hume's hopes vanish, "when I come to explain the principles, that unite our successive perceptions in our thought or consciousness" (T 635–636). He is using 'perceptions' as he does through all of paragraph eleven to mean those in the past train of perceptions under review. However, Hume is not suddenly just repeating that he can find no connection between the members of that train. He is rather lamenting a consequence of adhering to that earlier contention—namely, that he cannot sufficiently explain the connection in consciousness spread on them which makes us attribute to them simplicity and identity. Why not? Why couldn't he just give an analysis of the felt connection like the one I've suggested above? What more to the connection does there need to be?

In the next paragraph he tells us. He says there would be no difficulty "did our perceptions either inhere in something simple and individual, or did the mind perceive some real connexion among them..." (T 636). Unless someone (another or Hume later) comes up with a third alternative, one of these two is needed to resolve the difficulty. Now, if the relevant real connection—identity—were perceived, then the whole account of felt connection would be unnecessary. The explanation for the presence of an idea of the identity and simplicity of the many perceptions would simply be that they caused a faithful copy of themselves. However, this second alternative was rejected at the outset. So it is in the details of the first alternative that we can find an answer to the question, what more to the connection in consciousness does there need to be?

Think of the first alternative in terms of the core perceiver: the members of the past train of perceptions inhere in a simple, individual perceiver. Presumably, as Berkeley thought, the inherence of perceptions is just their being perceived.³⁵ Now, analyze the core perceiver into reflecting ideas with a felt connection. The inherence of the perceptions in the core perceiver, that is, their being perceived by it, is then just their being reflected—i.e., copied—in the reflecting ideas. Now what would it be for the core perceiver to be simple and individual? It would be for the reflecting ideas to be identical. Simplicity requires identity of parts. This goes along with what Hume says: he presents the first alternative—inherence in something simple and individual—as inconsistent with the principle that "all our distinct perceptions are distinct existences." So the first alternative is a case of *some* of our distinct perceptions *not being* distinct existences, namely, the reflecting ideas. So, whether or not at the level of the past perceptions, at the level of reflecting ideas the feared real connection—identity—is needed.

Thus Hume finds that a felt connection between perceptions in virtue of a felt connection between their reflecting ideas is not enough to explain the emergence of the idea of their simplicity and identity. He discovers that identity between distinct reflecting ideas is needed to explain the attribution of identity and simplicity to the many perceptions reflected upon. For the many perceptions to *seem* one, the ideas reflecting them have to *be* one.

That it is specifically the attribution of identity which requires a real connection between perceptions at some level can be supported by answering a possible objection. On my interpretation, if there is only a felt connection between the perceptions reflected upon, then there must be a real connection between the perceptions reflecting them. But, the objection goes, Hume has already considered and rejected this contention. In "Of the idea of necessary connexion" he emphasizes that the felt connection between cause and effect consists in the determination of the mind to pass from a perception of the cause to a perception of the effect (T 169). Then he gives the same account when we shift attention to the causal connection between the perceptions: that connection is just a felt connection consisting in the determination of the mind to pass from an idea of the perception of the cause to an idea of the perception of the effect. Presumably this regress would continue for as long as we were able to "change the point of view" to higher levels of ideas of ideas. Thus, it would seem, Hume explicitly rejects the contention that a merely felt connection at one level might somehow require a real connection at the next higher level.

I respond to the objection by pointing out a crucial difference between attributing causal connection to perceptions and attributing personal identity. In the case of causal connection, the two perceptions are taken to be *two*, whereas in the case of personal identity two or more perceptions felt to be connected are taken to be *one and the same*. Thus the crucial difference is that in the Appendix, and not in I iii 14, Hume is trying to explain "the principle of connexion, which...makes us attribute to them [distinct perceptions] *a real simplicity and identity*" (T 635). That a similar point confidently treated as a confirmation in I iii 14 was worriedly treated as a defect in the Appendix indicates that the difference between them is the explanation of the different reactions. That difference, then, is what explains the need for a real connection in the Appendix.³⁶

So, representing the many perceptions as a single thing on the basis of the felt connection requires a real connection between them or the ideas reflecting them. The relevant real connection is identity. So, representing the many perceptions *as* one and the same thing, requires that they or the many perceptions reflecting them *be* one and the same thing. Of course this is absurd. This is the inconsistency at the heart of Hume's worries.

It is not just an inconsistency, however; it is even worse. It is the affirmation of the identity and simplicity of the self, or at least of a core self.

Consider the two alternative "solutions" Hume gives for his difficulty: the first alternative is that the perceptions "inhere in something simple and individual." I have explained this alternative as their being perceived by the core perceiver, which boils down to their being reflected by ideas in consciousness. If these ideas are one and the same then the perceptions inhere in something simple and individual—a self—and so the self would be a unitary thing. It wouldn't matter if Hume were to say that the perceptions are separable from this simple, individual thing (T 207). It would still be a core self—a recalcitrant unitary thing. The second alternative solution is that the mind perceives a real connection between the perceptions we reflect on. If it did, then they would be one unitary thing. Additionally the ideas reflecting his identity would be identical, leading to the problem had by the first alternative. In either alternative the observing self is unitary; in the second the observed self is unitary too.

Of course, he is not simply affirming the unitariness of the self on either alternative. He is also denying it. That is the inconsistency. But this combination of affirmation and denial is what it is, Hume thinks, to endorse either of the then current theories of a unitary self. The first alternative—that endorsed by theologians—has it that distinct perceptions are identical with a single substance (T 43–44). The second alternative involves the mind discovering the truth of a theory like Shaftesbury's, that distinct perceptions are inseparable, so identical³⁷ (T 254). Hume was accustomed to being committed to absurdities, insofar as he was one of the vulgar. But it must have been galling to find himself committed as a philosopher to the same absurdities he derides in his opponents and thought he could avoid (T 232).

This is as much as can be gathered directly from the Appendix. The question remains: why would representing many perceptions as one require many perceptions that are one? The natural place to look for an answer is Hume's theory of representation.

IV. Representation and Misrepresentation

Hume's theory of how perceptions represent their objects is so taken for granted by him that he never explicitly lays it out nor argues for it. The relevant part of the theory is: representation requires resemblance, or at least some resemblance in salient respects. I suggest this is the source of his difficulty.

Hume notoriously leaves a gap in the Appendix. The source of the problem is left unstated. Every commentator has had to fill the gap somehow. It is a virtue in an interpretation to have the unstated source be something important that Hume often leaves implicit. His theory of representation by perceptions meets this criterion.

First I will describe the theory and show its role in the Appendix, then give textual evidence for it. I will focus on the representation by ideas, since the idea of self is at issue, but the same things will be true of representation by impressions.

The aspect of Hume's theory I am concerned with concerns the natural representation of ideas as opposed to their representation "beyond their nature" (T 20). This latter is the general representation an idea can acquire when used as an abstract idea (T 20–21). Even here resemblance in some small degree is required for representation. But greater resemblance holds between ideas and what they represent in themselves.³⁸ Consider perception insofar as it is an event in the brain. For Hume it is the presence of animal spirits in a certain region (also called "trace" or "cell") of the brain (T 60–61). Such a brain event "excites" an idea. Think of an idea as an image (T 1). There are senses of 'represent' in which the brain event represents the idea, and in which the brain event represents whatever the idea represents. I am concerned rather with representation by ideas. There are two sorts of objects represented by an idea: (i) what there is which the idea represents and (ii) what the idea represents there as being. Let me reserve 'intended object' for (i) and use 'intentional object' for (ii). For example, suppose someone mistakes a straight stick partially submerged for a bent stick. What there is which the idea represents—the intended object—is a straight stick. What the idea represents there as being—the intentional object—is a bent stick.

In the case of personal identity the intended object is several perceptions. The intentional object is a simple, identical thing. We represent many with an idea that represents there as being one.

Here is where Hume's account of representation gets him into trouble. For Hume an idea represents its *intended object* in part by resembling it. Ideas are copies of the things they represent, their intended objects. But an idea also represents its *intentional object* in part by resembling it. By means of its (salient) characteristics the idea represents there as being something with those characteristics.³⁹ So to represent what *are* many as one (and not many), would require an idea that is many and not many.

This is why a real connection, specifically identity, is required in the Appendix. For the many past perceptions to be represented as one, there are two alternatives. They could be many with a discoverable real connection—identity—between them. In that case, an idea reflecting them would represent the many as one. Alternatively it could be that the past perceptions are many but the many ideas reflecting them have a real connection, i.e., are identical. So the reflecting ideas would be one and could represent their objects as one. Unfortunately each alternative is inconsistent.

There is a good deal of mostly indirect evidence that Hume held this resemblance theory of representation. Consider first the evidence that ideas resemble their *intended objects*. The most direct is where Hume baldly asks,

"For how can an impression represent a substance, otherwise than by resembling it?" (T 233). And again, "A substance is entirely different from a perception. We have, therefore, no idea of substance" (T 234). In another place he says, in a subordinate clause, "an image necessarily resembles its object" (T 260). In "Of the origin of our ideas" Hume says ideas are "images" of impressions (T 3). There is a "great resemblance" between them (T 2). An idea is "in a manner the reflexion" of the impression. Ideas of his chamber are "exact representations" of the impressions of it (T 3). He says "all simple ideas and impressions resemble each other" (T 4). And he concludes "that ideas are preceded by other more lively perceptions, from which they are derived, and which they represent" (T 7). Later he says ideas are "copies and representations" (T 19). Much later he says, "All ideas are deriv'd from, and represent impressions" (T 161). The ideas represent the impressions they copy. Hume says that a perception of a succession of notes is a succession of perceptions (T 36–37). And he says that such a succession of perceptions can only by a fiction be taken to represent something that is not a succession—an unchangeable object—and thus which does not resemble them in the relevant respect. The fiction is that the object is a succession and thus resembles the idea. He thinks impressions are the sort of thing that easily can be taken for a hat, or shoe, or stone (i.e., the things the impressions represent) (T 202). And he emphasizes that an impression and an object ("external existence") can have a quality in common. A quality of an object "cou'd not be conceiv'd, unless it were common to an impression" (T 242). He says that an idea of extension (i.e., of extended things) is itself extended (T 239–240). In 'Of personal identity' he assumes that an impression of the self "must continue invariably the same, thro' the whole course of our lives; since self is suppos'd to exist after that manner" (T 251). All these examples suggest that Hume thinks perceptions resemble their intended objects. He also thinks this is the standard philosophical view: philosophers "distinguish betwixt the objects and perceptions of the senses; which they suppose co-existent and resembling..."⁴⁰ (T 202).

Cases of fictions show that he thinks ideas resemble their *intentional objects*. In "Of the origin of our ideas" Hume has told us that simple ideas resemble their intended objects. Fictions are complexes of simple ideas. For example an idea of an apple consists of simpler ideas of the taste, feel, shape, color of an apple (T 2; see also T 221). Thus it consists of ideas that *have* these characteristics. Replace the idea of red with the idea of the color of the metal gold (which idea would have that color) and the result is an idea of a golden apple. There is no intended object of the idea. The intentional object is determined by and so has the relevant characteristics of, the simple ideas in the complex. It is presumably by this process that Hume is able to imagine to himself "such a city as the New Jerusalem, whose pavement is gold and walls are rubies, tho' I never saw any such" (T 3). Another example

is the case in which the senses “represent as minute and un compounded what is really great and compos’d of a vast number of parts” (T 28). With minute ideas (than which “nothing can be more minute”) we represent things as minute. Thirdly, Hume says that an idea of a double existence—that is of impression and distinct object—cannot simply be a copy of a single impression of the senses. Because it is single it cannot represent there as being two things (T 189).

So Hume thinks, I suggest, that representation of both intended object and intentional object is accomplished at least in part by resemblance. Furthermore, the fact that the theory is almost always left implicit suggests that Hume often thought it was too obvious to bother mentioning. This would account for the fact that he does not mention it in the Appendix, even though it is the source of his difficulty.

I have said that the source of the problem is Hume’s theory of representation. I might more perspicuously say it is his theory of *misrepresentation*. One might worry at first that he has no such theory. How would an idea *misrepresent* something, say, blue as, say, red? To have a blue intended object the idea must be blue; to have a red intentional object the idea must be red and so not blue. So the idea must be blue and not blue. The answer is that one represents the intended object via different sorts of characteristics than the one the object is misrepresented as having. Consider your image in a fun-house mirror. You know it is an image of you because of the relative locations of the colors. But the image by being oddly shaped represents you as being oddly shaped. The image represents its intended object by resembling in some respects, and *misrepresents* it by *misresembling* in some other.⁴¹

Unfortunately for Hume, this theory of misrepresentation does not work for misrepresenting the many past perceptions as a single thing. The ideas with which they are remembered are memories. The memory “preserves the original form, in which its objects were presented...” (T 9). So any ideas that resemble in enough respects to be memories will thereby represent the train of past perceptions as being many, so cannot misrepresent the train as being one.

To be sure, memory can be imperfect and give inadequate reflections of what went before. Hume has seen Paris but cannot perfectly represent all its streets (T 3). Nonetheless it is clear from Hume’s discussion that the multiplicity of the past train of perceptions is too manifest. The person reflecting on them fights his tendency to consider them a steadfast object, invariable and uninterrupted. That is, he fights his tendency to consider them as one. But he cannot, and yields, and asserts “that these different related objects are in effect the same, however interrupted and variable”⁴² (T 254). The result is representing them as many *and* as one. Yet, as shown above, the same ideas cannot represent them as many and as one.

It might seem that Hume has an alternate theory of misrepresentation available that will let the same ideas do both: to represent many as one and the same is simply to represent them with many ideas that have such a close felt connection that the mind overlooks the transitions (T 220). But this alternate theory won't work. Hume explicitly says such an idea "to an accurate view affords as perfect a notion of *diversity*, as if there was no manner of relation among the objects" (T 253). He makes clear that "these two ideas of identity, and a succession of related objects [are] in themselves perfectly distinct, and even contrary..." (T 253). And he explains their being "confounded" as substituting one for the other. Resemblance makes us "substitute the notion of identity, instead of that of related objects" (T 254). What we do is stop representing the related perceptions as related perceptions and begin representing them as identical. The mind's overlooking the transitions just is this idea coming to be substituted. We do this even "tho' we incessantly correct ourselves by reflexion, and return to a more accurate method of thinking"—i.e., return to thinking of them as a succession of related objects (T 254). Finally we "yield" to the "bias" toward attributing identity. This alternation could not take place if representing a related succession as identical were nothing but representing it as a related succession.

Thus, we must conclude that the very memories themselves which represent a past train of related perceptions cannot additionally represent them as identical. This is so no matter how closely connected the memories feel.

There is, however, a second major way to misrepresent in the *Treatise*. The first was to have an idea that fails to resemble in some respects its intended object. The second is to confound ideas. As just mentioned, to *confound* is to substitute for an idea of something, a related idea of something else. As Hume says, "For we may establish it as a general maxim in this science of human nature, that wherever there is a close relation betwixt two ideas, the mind is very apt to mistake them, and in all its discourses and reasonings to use the one for the other" (T 60). The relevant relations are those familiar from the principles of the association of ideas—resemblance, contiguity, and causation. The mind often presents an idea related in these ways "in lieu of that which the mind desir'd at first to survey." We generally don't notice the substituting but proceed with our reasoning as if we were using the same idea (T 61, 203). In addition to relations between ideas, resemblance of the actions of the mind in considering ideas can lead to the inadvertent substitution. "[W]e may in general observe, that wherever the actions of the mind in forming any two ideas are the same or resembling, we are very apt to confound these ideas, and take the one for the other" (T 61, T 205, note).

The resemblance in the actions of the mind is precisely what occurs in the case of attributing identity to a succession of related objects (T 253–254). Thus it might seem that Hume's account of confounding is exactly suited to explaining how we represent many as one: it is simply that an idea of one is inadvertently substituted for an idea of many.⁴³ If this is so then my interpretation of the Appendix would fall apart. I have claimed that in the Appendix Hume finds that he cannot explain how we misrepresent many past perceptions as one single thing. If the problem is easily and obviously solved by the mechanism of confounding, then I am wrong that it is a problem for Hume and therefore wrong that it is the problem baffling him in the Appendix.

The key to this possible solution would be to have the ideas that accurately reflect the past perceptions be different from the idea that represents there as being one simple thing. Inadvertently substituting the latter idea for the former would then be misrepresenting the past perceptions as the simple thing. So at first there is an accurate train of memories which represents the past perceptions as many. The question is, what idea is substituted and how does it represent them as one?

It might seem possible that we just inadvertently substitute an idea of an arbitrary single thing. The only claim this idea would have to being of the many past perceptions, would be this: that it was substituted inadvertently for the train of ideas of them, and used in our reasoning as if there had been no substitution.

But this is not what happens in attributing identity to the self. We become very aware of the substitution and struggle against it before yielding to a recognized absurdity "that these different objects are in effect the same, however interrupted and variable" (T 254). Thus, we do not unawares shift our thinking to something else. Or at least we become aware and then continue to think of the same things as before. The substituted idea of identity continues to have the many past perceptions as its intended objects. So the question remains, what idea is substituted and how does it represent them as one?

Hume himself gives an example that may help. Sometimes the senses "represent as minute and uncompounded what is really great and compos'd of a vast number of parts" (T 28). In these cases because of distance or lack of acuity, the eye conveys a single simple impression to represent a multitude in the world. For example a mite with its many parts appears as the tiniest speck. This seems to be a case of representing many as one.

Use of this example is complicated, however, by Hume's outright assertion that what is "a compound object, can never be represented by a simple impression" (T 231). Hume seems to be contradicting here what he said about the mite. The contradiction can be resolved by distinguishing an "adequate idea" (alternatively, "just notion") from one that isn't. An

adequate idea of something has distinct component ideas representing every part of that thing (T 28). An inadequate idea doesn't. Thus a compound object cannot be represented as one by a simple *adequate* idea. However it can be by a simple *inadequate* idea. What holds for ideas holds also for impressions, so a compound object can never be represented by an *adequate* simple impression. This must be what Hume meant at T 231.

So an inadequate idea can represent a many as one. The next question is, how does it represent its intended many? There must be a salient respect in which the unitary idea resembles the intended multiplicity. For Hume's example such a respect is not hard to give: suppose the relevant simple idea is visual. Its position in the visual field—i.e., in the current assemblage of visual impressions—resembles closely the position of its intended complex object in the real scene represented by the visual field. The inadequate simple idea resembles its intended complex object with respect to relative place.

A similar account could be given for how a unitary idea would represent a past succession of perceptions. It would represent them as one by being one. What is needed is an appropriate respect of resemblance to make the succession of perceptions be the intended object. Such a respect might be the "je-ne-scai-quoi" Hume mentions at T 106—something we conceive of "in thinking of our past thoughts," i.e., "the action of the mind in the meditation." The remembered perceptions may be various, but all would resemble an idea of them in this respect.⁴⁴

Thus it seems that one's past perceptions could be represented as identical by inadvertently substituting an inadequate idea of them as one for an adequate idea of them as many.

There is a problem with this proposal however—related to the problem with the previous proposal. Hume's description of acquiring the idea of the self involves a stage in which we attribute unitariness to what we fully recognize are manifold (T 254). It is not sufficient that the substituted idea merely *have* the same intended objects (viz., the many past perceptions) as the train of ideas substituted for. We must *recognize* the sameness, when having the ideas.

In other words, it is not enough that the intended objects of the many reflecting ideas be the same as the intended objects of the single inadequate idea. They must be represented as being the same. That is, the simple inadequate idea must represent them as being something which is the same thing as what the many reflecting ideas represent them as being. But the many ideas represent there as being many in succession. The single inadequate idea represents there as being one, steadfast and uncomposed. How can we represent these many as being the same as this one?⁴⁵ We would have to be able to represent there as being many distinct things that are identical with each other. But this was the original problem. The

proposed solution in terms of substituting ideas has the same problem it was meant to resolve.

To put it briefly: the problem is to explain how we represent a recognized many as one. The proposal is that we represent it first as many, then as one, while recognizing that we are representing the same thing(s). But you cannot explain this unless you can explain how we represent a recognized many as one. The original problem recurs in the solution.⁴⁶

So the solution based on the mechanism of confounding begs the question. It might seem that appeal to Hume's resemblance theory of representation can help. The intentional object of an idea is determined by the idea's salient characteristics. So perhaps ideas with some of the same salient characteristics can represent their intentional objects as numerically the same. Then if the many reflecting ideas and the single inadequate idea resemble, perhaps they represent there as being a many the same as the one. But this is hopeless. There is no way to distinguish between representing a many that is one larger than the original many and representing the many as the one. Resemblance is not sufficiently fine-grained to determine how many things you are representing there to be, or which are represented as numerically the same.⁴⁷

The need is for a way to represent that which is represented as many as the same as that which is represented as one. This suggests a last resource in the *Treatise* for explaining the misrepresentation—the account of the “principle of identity.” There Hume explains how we misrepresent a single unchanging thing alternatively as one thing and as many things in succession. We do this by alternating “views” of it (T 201). Since the problem is with attributing identity, Hume's account of the idea of identity is a natural place to look for a possible solution. Further, this discussion of identity is needed to make my account more accurate. I have talked of representing many as one. But this is just half of representing distinct things as identical. One must represent them as one, on one view, and as many, on another.⁴⁸

Even if this appeal to Hume's account of the principle of identity were to help, however, it wouldn't completely solve the problem given in the Appendix. That problem is also a problem with attributing simplicity, and Hume does not have an account for simplicity parallel to that for identity. He never says a simple thing is one from one view and many from another. So already this possible solution is incomplete: it cannot solve the problem with attributing simplicity. To that extent, the Appendix problem would remain unaddressed.

In any event, the proposal is that by alternating ideas of something as one and something as many we can misrepresent it as both one and many, and as the same thing either way. But this proposal is of no help. Alternating

ideas is just successive substitution of one for the other. The appeal to substitution to solve the problem has already been discredited.

So the appeals to confounding and to the idea of identity have been in vain. The difficulty remains. Hume cannot explain how to represent a recognized many as one.

V. Conclusion

There are three important criteria, generally agreed upon by commentators, for judging the success of an interpretation of Hume's Appendix.⁴⁹ First, can it make sense of Hume's maddening apparent claim that two consistent principles are inconsistent? Second, does it explain why he says what he says and why he doesn't say what he doesn't say? Third, does it explain why the difficulty comes up only in the course of considering the self and the idea of the self? My interpretation answers these questions as follows:

First criterion: notoriously Hume says,

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 connexion among distinct existences.* (T 636)

It has never been clear why Hume thinks these are inconsistent, nor what he would gain by renouncing one or the other of them. On my interpretation the inconsistency is this: we represent many distinct perceptions as one and the same thing. Therefore we have an idea of them that is many distinct ideas that are somehow the same idea. We could get this mongrel idea either by (i) discovering an identity among distinct existences or by (ii) having the distinct ideas reflecting them be identical. But (i) is inconsistent with the second principle, while (ii) is inconsistent with the first. So Hume has to deny one or the other of his principles. Given his account of acquiring the idea of the self, he can't consistently hold both.

Second criterion: every word Hume uses in the Appendix is explained by my interpretation. What he leaves out is also explained: it is something he almost always leaves implicit—his theory of representation. Filling the gap in the Appendix with something Hume almost always leaves implicit explains the presence of the gap.

Third criterion: why did this problem come up only concerning the idea of personal identity? A general problem such as I have proposed would seemingly have occurred to Hume when discussing other "instances of this tendency of relation to make us ascribe an *identity to different objects*" (T 204). Examples are his discussions of the idea of body in I iv 2, "Of

scepticism with regard to the senses," and of the fantasies of the ancient philosophers in I iv 3, "Of the antient philosophy."

However the question ought not presuppose that on a correct interpretation the Appendix difficulty could not have come up elsewhere. That would be to presuppose without argument that the difficulty is not a general one.

If the question remains simply, why *didn't* the difficulty come up in other contexts, a perfectly fair answer is: Hume just didn't think of it. After all, how else would an interpretation explain the afterthought aspect of the Appendix difficulty? It is appropriate to wonder, why didn't Hume discover the difficulty while writing I iv 6? The best answer would seem to be that he just didn't think of it. What's sauce for I iv 6 is sauce for the other contexts.

Nonetheless I think an interpretation should explain why reviewing I iv 6 was a *more natural* way for Hume to see the difficulty, even if it is not the only possible way. So I will speculate why he was less likely to see it in other contexts.

In I iv 3 Hume's discussion of attributing simplicity to complex objects depends heavily on his discussion of attributing identity to successions. The latter depends heavily on his I iv 2 discussion of attributing identity. So if he didn't see any problem in I iv 2, it is unlikely he would see one in I iv 3. In I iv 2 the main discussion concerns attributing continued existence because of *constancy*: the resemblance of successive interrupted impressions (T 199). This resemblance would have made problems with representation harder to see. Representing distinct things in succession as identical would have involved alternating many ideas of them which represent them as many, with one idea of them which represents them as one. The resemblance between the successive things might have seemed to allow a single idea to represent all of them. It might have made it seem easy for the inadequate idea representing them as one to have the same intended objects as the adequate ideas representing them as many. Of course this whole account depends heavily on the discussion of the principle of identity (T 200–201). His assuming that there is no problem there would have allowed Hume to see no problem in discussions that depend on it. And again, the resemblance of all the perceptions involved might have disguised the problem. However, as I have argued, there is a problem—one the resemblance of all the perceptions involved does not mitigate. Had Hume subjected this passage to "a more strict review" (T 633), he could well have discovered there the basic problem of representing many distinct things as one and the same.

But he didn't. He subjected the account in I iv 6 to that review. This is a natural place for Hume to want to be extra careful. After all his "science of MAN" (T xv) is mainly a science of the soul or mind. (He doesn't distinguish them.) The elements of that science are perceptions; the gentle forces he posits between them and the other principles he posits govern the

co-occurrence, sequence, and transfer of vivacity between perceptions. In giving this theory he hopes to make us more "thoroughly acquainted with the extent and force of human understanding" and more able to "explain the nature of the ideas we employ, and of the operations we perform in our reasoning" (T xv). Hume is following Descartes here in assuming that "MAN" (and by this I take him to be referring to people in general) is a thinking thing. Hume goes on to show that, more accurately understood, a person is a feeling thing. In any event, for Hume human nature is the nature of our minds. He also assumes with Descartes that the idea of one's self is the idea of one's mind (T xvii–xviii). So a problem with the idea of the self is a problem with the idea of the object of his science.

Thus Hume was amply motivated to more strictly review I iv 6. Now there, as everywhere else, he appeals to the conceit of the core perceiver transiting the related ideas. The mind of the core perceiver played the role of a simple substance in which perceptions inhered. So given the conceit, there was no problem. Hume was apparently confident that he could make use of this conceit as long as it was clear that ultimately the mind of the core perceiver is analyzable into reflecting ideas attracting others to consciousness. But only in the context of characterizing the self was he forced to think through the implications of this analysis. Leaving the core perceiver unanalyzed would be leaving a core unitary self. Analyzing it led to the realization that he could not give a consistent account of the idea of the self.

Were the perceptions composing the self fairly constant, Hume might have seen no problem even here. But the self involves quite different perceptions. Think of Hume's poor fellow dragged from agreeable garden and company to terrifying dungeon (T 245). This disparity in perceptions makes the question more pressing how an idea inadequate enough to represent them as one could be adequate enough to be representing *them*.

Thus although Hume could have discovered the general problem by review of the principle of identity passage, it is more natural that he discovered it the way he did—by review of I iv 6.

Even granting this answer, however, a related question arises. If the problem discovered by reviewing I iv 6 is as general as I claim, why didn't Hume note that generality in the Appendix? Why does he confine his remarks to the idea of personal identity?⁵⁰

I think that the way he formulated the problem to himself obscured its generality. As I have argued, Hume worries that the reflecting ideas have to be connected into a simple, identical self in order to represent the past perceptions as being connected into a simple, identical self. Putting it this way keeps the focus on the self. I've argued that this worry makes sense only if *at root* it is a worry about representing a recognized many as one. Were he to have formulated the root problem clearly, he would likely have remarked on its generality. But he didn't formulate it that way. He wasn't clear about

it. If he had been clear as to what the problem was, he would have made it clear to us.

I am assuming that even the greatest philosophers can sense a problem before they are in a position to fully formulate it. We as participants in a group effort to understand them can in principle at least sometimes give better formulations. This is so even if we, without their insights, would ourselves have little to say.

Suppose however that I am wrong about which problem it is that Hume is grappling with in the Appendix. Nonetheless, I have identified a difficulty for Hume which to all appearances he cannot resolve. In this respect my interpretation is on a par with almost all—and perhaps all—of its main competitors.

The consequences of the difficulty are drastic. One of Hume's main concerns in the first book of the *Treatise* is to answer the question, "Why do we believe what we do?" He supplies answers for some of our most basic beliefs—that there is necessary connection, that there are bodies, that there is a unitary self.⁵¹ His general strategy toward our basic beliefs is two-fold. First, he argues that the belief is unreasonable; there is either no reason to believe it or there is reason not to believe it. In this he is a skeptic. Second, he gives a causal explanation for the genesis of the belief based on certain assumptions about the natural operations of the mind. In this he is a mental naturalist. Hume's opponents are those who think explanations of the operations of the mind are paradigmatically normative—we do what we do because we are trying to do what we ought. In the case of beliefs, the norms of reason govern what we ought to believe. Hume's naturalistic explanation leaves out any appeal to norms, goals, or intentions—we do what we do because that is what things like us tend to do. His skeptical arguments are meant to refute his opponents by showing that none of our basic beliefs are what we ought reasonably to believe. His naturalistic explanations then become the replacement answer to the question "Why do we believe what we do?"⁵²

In the case of belief in the self this two-fold strategy was supposed to work as follows. It is unreasonable to believe in a unitary self because the self is clearly just many perceptions. Nonetheless certain relations between the perceptions cause ideas of them naturally to introduce each other, which causes an idea of the identity of the perceptions, which idea then becomes enlivened. However Hume discovers in the Appendix that the strategy subverts itself. Only if the many ideas representing the many perceptions are somehow identical can they represent the perceptions *as* identical. But many ideas, somehow identical, of a variety of perceptions, would count for Hume as the sort of unitary self he was at pains to discredit. So Hume finds that the self cannot both be many and represent itself as being one. Concerning the

basic belief in a unitary self, Hume cannot both be a skeptic who refutes it and a naturalist who explains its genesis. If, as I have suggested, the root problem is more general, Hume is similarly torn concerning the basic belief in body.⁵³

Hume's only recourse is to "plead the privilege of a sceptic, and confess, that this difficulty is too hard for my understanding" (T 636). He is a skeptic, it's true, but enjoying this privilege is cold comfort here. His skeptical arguments are mostly means of clearing the way for his science of man, of human nature. By "the application of experimental philosophy to moral subjects" (T xvi) he hoped to achieve even greater glory for himself and his native country than Newton had by its application to "natural" subjects. Hume's procedure was to be a determining by "careful and exact experiments" of the "powers and qualities" of the mind (T xvii). The mind is the object of Hume's science. An irresolvable contradiction concerning the mind threatens that science at the core. All Hume can do is hope that the difficulty is not "absolutely insuperable," that someone "may discover some hypothesis, that will reconcile these contradictions" (T 636).

NOTES

I am grateful for suggestions or criticism from Mark Rubin, Elise Springer, Eddie Zemach, Jerome Shaffer, Phillip Cummins, Wade Robison, Martha Nussbaum, James van Cleve, John Wright, Sandra Baxter, some anonymous referees, and Don Garrett.

1 Robert J. Fogelin, in *Hume's Skepticism in the Treatise of Human Nature* (London: Routledge and Kegan Paul, 1985), 100-108, rightly points out the general lack of textual support in the Appendix for most interpretations. I fear, however, that he doesn't show that this criticism wouldn't apply to his own proposals. Hereafter, "Fogelin."

2 Barry Stroud, in *Hume* (London: Routledge and Kegan Paul, 1977), 133-134, makes a related point, as have others. A concern with the idea is in line with Hume's "philosophy, which pretends only to explain the nature and causes of our perceptions, or impressions and ideas" (T 64). Hereafter, "Stroud."

3 David Hume, *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd edition revised by P. H. Nidditch (Oxford: Clarendon, 1978). Hereafter abbreviated "T," with page numbers inserted parenthetically in the text.

4 A. H. Basson, *David Hume* (London: Penguin, 1958), hereafter, "Basson"; Tom L. Beauchamp, "Self Inconsistency or Mere Self Perplexity," *Hume Studies* 5 (1979): 37-44; Daniel Flage, *David Hume's Theory of Mind* (London: Routledge, 1990), hereafter, "Flage"; Don Garrett, "Hume's Self-Doubts about

Personal Identity," *The Philosophical Review* 90 (1981): 337-358, hereafter, "Garrett"; H. P. Grice and John Haugeland, "Hume on Personal Identity," unpublished manuscript; Louis E. Loeb, "Causation, Extrinsic Relations, and Hume's Second Thoughts about Personal Identity," *Hume Studies* 18.2 (1992): 219-231; D.G.C. MacNabb, *David Hume: His Theory of Knowledge and Morality* (London: Hutchinson, 1951), hereafter, "MacNabb"; S. C. Patten, "Hume's Bundles, Self-Consciousness, and Kant," *Hume Studies* 2.1 (1976): 59-64; David Pears, "Hume on Personal Identity," *Hume Studies* 19.2 (1993): 289-299; Wade L. Robison, "In Defense of Hume's Appendix," in *McGill Hume Studies*, edited by David Fate Norton, Nicholas Capaldi, and Wade L. Robison (San Diego: Austin Hill Press, 1976), hereafter, "Robison, 'Defense'."

5 Oliver Johnson, *The Mind of David Hume* (Urbana and Chicago: University of Illinois Press, 1995); Norman Kemp Smith, *The Philosophy of David Hume* (London: Macmillan, 1941), hereafter, "Kemp Smith"; John A. Passmore, *Hume's Intentions* (Cambridge: Cambridge University Press, 1952); H. H. Price, *Hume's Theory of the External World* (Oxford: Clarendon, 1940); Wade L. Robison, "Hume on Personal Identity," *Journal of the History of Philosophy* 12 (1974): 181-193, hereafter, "Robison, 'Personal Identity'." These interpretations are applications of Kemp Smith's larger contention that Hume inconsistently holds under Newtonian influence that the self is only a bundle of perceptions, and yet under Hutchesonian influence that there is a self that operates as an "ever-present observer" (Kemp Smith, 73). See also related worries in MacNabb (151-152) and Terence Penelhum, "Hume's Theory of the Self Revisited," *Dialogue* 14 (1975), 406-407, hereafter, "Penelhum, 'Revisited'."

6 Refutations of most of the interpretations cited so far are found in Garrett, Fogelin (100-105), and Penelhum ("Revisited"). Garrett's and Penelhum's articles contain refutations of Kemp Smith's claim that the account of self in Book I is incompatible with the awareness of self required in the Book II accounts of sympathy and of the indirect passions. It seems to me that were the Book I account of the fictional idea of the self successful, that idea, sufficiently enlivened, could play any role required by Book II. This much can be said for Kemp Smith's position, however. It may be that the relevant awarenesses of the self in Book II require a lively idea of the self as unitary. If so, then Kemp Smith would be right that the Book II appeals to the idea of self are inconsistent with the Book I account of the self as many. He would be right, that is, if I am right that Hume can't explain how an idea could represent the many perceptions as something simple and identical.

7 For example see Nelson Pike, "Hume's Bundle Theory of the Self," *American Philosophical Quarterly* 4 (1967): 159-165; Penelhum, "Revisited," 405-406; Stroud (131), and Garrett. Presupposing that there must be further explanation is the mistake behind Nathanson's proposal in "Hume's Second Thoughts on the Self," *Hume Studies* 2 (1976): 36-45, that the mind for Hume is a "set of dispositions" (40) or a "possessor of propensities" (45); behind Waxman's unargued claim that association requires "retentive memory" (236-237); and behind Jane L. McIntyre's contention, in "Is Hume's Self Consistent?," in *McGill Hume Studies*, edited by David Fate Norton, Nicholas Capaldi, and Wade L. Robison (San Diego: Austin Hill Press, 1976), that past

perceptions have to overlap with present perceptions in order to affect them. McIntyre's account is admirable, though, in its attempt to address the concern of the second group of interpretations while staying true to the bundle theory of the first.

8 Corliss Gayda Swain, in "Being Sure of One's Self: Hume on Personal Identity," *Hume Studies* 17.2 (1991): 107-124, has relied on such interpretations to argue that Hume in the Appendix is not dissatisfied with his theory of personal identity (107; hereafter, "Swain"). However, to go on to say his dissatisfaction with his "former opinions" concerns nothing in I iv 6, but rather his pre-philosophical opinions that there can be perfect identity through change, is stretching the text too far (Swain, 117).

9 L. Ashley and M. Stack, "Hume's Theory of the Self and Its Memory," *Dialogue* (1974): 239-254; John I. Biro, "Hume on Self-Identity and Memory," *The Review of Metaphysics* 30 (1976): 19-38; Kemp Smith; John Laird, *Hume's Philosophy of Human Nature* (New York: E. P. Dutton, 1931); MacNabb; Terence Penelhum, "Hume on Personal Identity," *The Philosophical Review* 64 (1955): 571-589 and "Revisited"; Swain.

10 The only way I know of to make sense of the identity of distinct things is the theory I propose in "Many-One Identity," (*Philosophical Papers* 17 [1988]: 193-216) and "Identity in the Loose and Popular Sense," (*Mind* 97 [1988]: 575-582). But such a theory can't in good conscience be attributed to Hume or to these commentators.

11 See T 256 and T 260.

12 Hume's account in this respect is just like his account of getting the idea of body: "[W]e are not apt to regard these interrupted perceptions as different, (which they really are) but on the contrary consider them as individually the same..." (T 199).

13 It is true that Hume seems also to talk about the fictitious identity of bundles to themselves. He does seem also to think that having a succession of parts is contrary to something's identity through time (T 255-258). He thinks this, I would guess, because he thinks that strictly speaking nothing with parts exists. Only the parts do (T 30). So the problematic talk of a succession lacking identity boils down, for Hume, to the unproblematic claim that the members of the succession are distinct. In any event the main points of I iv 6 relevant to understanding the Appendix could be made while dispensing with any apparent denials of the identity of successions with themselves, instead staying strictly to denials of the identity of members of a succession with their successors.

I still haven't done justice to Penelhum's discussions in particular. All I've done is to locate, I think, where Penelhum's fundamental disagreement with Hume lies: it is the question whether bundles exist. Hume's denial that they do would lead him, I conjecture, to deny something that Penelhum assumes—that number is relative to concept. (For more on the denial of the existence of wholes with parts see my "Hume on Infinite Divisibility," *History of Philosophy Quarterly* 5 [1988], 135-137, and "Corporeal Substances and True Unities," *Studia Leibnitiana* 27 [1995]: 162-167.)

14 This fact is reinforced by the fact that Hume is also trying to explain the

attribution of simplicity to what is really complex (T 635).

15 "'Tis the composition of these, therefore, which forms the self" (T 634).

16 See my "A Defense of Hume on Identity through Time," *Hume Studies* 13 (1987): 323-342, for further discussion.

17 See also Hume's endorsement of Malezieu's argument that anything with parts is only fictitiously one (T 30).

18 In the *Abstract* of the *Treatise* he uses the phrase "without any perfect simplicity or identity" (T 657).

19 Flage, 147-148.

20 Hume's third paragraph also provides a quick rebuttal to the position Flage attributes to Shaftesbury that the self is composed of inseparably connected distinct perceptions. Hume holds if they are distinct then they are separable, which would make Shaftesbury's position inconsistent (T 634; see also Flage, 83-87).

21 George Berkeley, *A Treatise Concerning the Principles of Human Knowledge*, in *The Works of George Berkeley*, edited by A. A. Luce and T. E. Jessop, Volume 2 (London: Thomas Nelson, 1948-1957), Section 20. Hereafter, "Berkeley."

22 He asks, "What then gives us so great a propensity to ascribe an identity to these successive perceptions...?" (T 253). He proceeds to show that "the identity, which we ascribe to the mind of man, is only a fictitious one..." (T 259). To falsely ascribe identity to some things is to have a fictitious idea of them as identical which is subject to enlivening.

23 It is not strictly a principle because Hume derives it from two other unargued claims: that the distinct are conceivably separable (T 18; see also T 10, 36, 38, 79, 233) and the conceivable is possible (T 19-20). Bricke calls the first of these "Hume's 'separability principle'" (68). For more on separability and inseparability in Hume see my "Abstraction, Inseparability, and Identity," in *Philosophy and Phenomenological Research* 57 (1997): 307-330.

24 John Locke, *An Essay concerning Human Understanding*, edited by Peter H. Nidditch (Oxford: Clarendon, 1975), II xxvii 9. Hereafter, "Locke."

25 Locke, II xxiii 2. See also Berkeley, Part I, Section 1.

26 Giving this up would undermine his skeptical attacks on our belief in necessary connection and in body. That the former would be undermined is obvious from his discussion at T 155-167. The latter would be undermined if the identity which we "feign" (T 208) were real and discoverable (T 199-210).

27 "When we reflect on our past sentiments and affections, our thought is a faithful mirror, and copies its objects truly...." (*An Enquiry concerning the Principles of Morals* in David Hume, *Enquiries concerning Human Understanding and concerning the Principles of Morals*, edited by L. A. Selby-Bigge, 3rd edition revised by P. H. Nidditch (Oxford: Clarendon, 1975), 17-18. Hereafter, references to the first *Enquiry* abbreviated as "EHU"; references to the second *Enquiry* abbreviated "EPM."

28 The passage is purportedly about coming to *believe* that many perceptions are a simple, identical thing. Actually, there he doesn't explain the belief, per se. No mention is made of the source of the vivacity of the

idea. But doubtless we are to assume his account would be as in I iv 2: the belief in continued and distinct objects borrows its vivacity from the impressions or the impressions vividly remembered, mistakenly taken to be identical (T 208). Likewise the vividly remembered perceptions taken to be the self would lend vivacity to the idea of the self. Even ideas can be vividly remembered when "not...consider'd, as the representation of any absent object, but as a real perception in the mind..." (T 106).

29 This is a common theme. Hume varies among 'imagination', 'thought', 'mind', and 'we'. I will uniformly use 'mind'. See examples of the varied expressions at T 202, 203, 204 220, 254, 260, and 635.

30 See especially his account of acquiring the idea of body at T 200–205. An example of the sort of thing Hume is talking about can be readily seen by making a flip movie. Put dots in slightly different positions on successive pages of an old book, then riffle the pages. There appears to be a single jittery dot. That dot is really the several dots in succession falsely believed to be identical with each other.

31 Why would Hume employ this conceit at all in the context of giving his bundle theory of mind? My guess is that he is illustrating the force of the natural belief in a unitary self by retaining it even in the course of showing it to be false.

32 Bricke is clear about this distinction between observer and observed, though the problem he adduces on 88 is very different from the one I am giving (Bricke, 83–88). In presenting his problem, he assumes that the present train of reflecting ideas must temporally overlap the past train of perceptions. As Bricke maintains in note 13, however, Hume does not hold this assumption in the Appendix (Bricke, 159). Bricke thinks denying it brings another problem to Hume: how are one's present perceptions represented in the idea of the self? This apparently assumes that an idea of the self has to be of all the perceptions that compose the self. But why? An idea of Paris need not be of all the streets composing Paris (T 3). Bricke could grant this rebuttal of his assumption, yet still raise a related question: why do we think that present perceptions belong to the self? I would guess that the answer is something like induction. Any time we reflect on perceptions we get the idea of the self. This makes us think every perception is dependent on a self, even those we haven't reflected on (T 222).

33 Reflecting ideas are copies of the perceptions they reflect, just like mirror images are copies of what they reflect. This has nothing to do with what Hume calls "impressions of reflexion" and the ideas copied from them. Such impressions are sentiments caused by ideas of our impressions of sensation (T 7–8). I thought of using the term 'apperception' where I am using 'reflecting idea' to denote ideas of perceptions. But this term has no precedent in Hume and may remind the reader too strongly of Leibniz or Kant. The impressions of smoothness and ease are, on the other hand, impressions of reflexion.

34 An example of the former is Butler ("Of Personal Identity," first appendix, in *The Analogy of Religion*, edited by Samuel Halifax [Oxford: Oxford University Press, 1849]); an example of the latter is Locke, II xxvii 10.

35 Berkeley, Part I, Section 49.

36 See T 174: "The difference in the effects of two resembling objects must proceed from that particular, in which they differ."

37 The fact that both the theologians and Shaftesbury posit something hidden to disguise the contradiction, (T 254–255) does not change the fact that both are committed (according to Hume) to the contradiction of thinking many perceptions are also somehow identical.

38 Flage gives an example of the representation of causes by means of ideas of their effects. This would seem to be a counterexample to the resemblance claim. Here again, though, the idea would be representing beyond its nature (Flage, 54–55).

39 These have to be characteristics that can be recognized by attention to the idea. Also, normally, characteristics an idea has just insofar as it is an idea do not count. The idea's having the salient characteristics it does is what makes it have a resembling intentional object.

It might be that this account of representation can explain how the mind manages to "spread itself on external objects" (T 167). Perhaps the mind mistakenly treats a characteristic of an accompanying impression as a salient characteristic of a given idea, thus injecting the characteristic of the impression into the intentional object of the idea. The intended object of the idea would then be misrepresented as having the characteristic. I am indebted to James van Cleve for provoking this line of thought.

40 As is shown also in the first *Enquiry*, Hume takes this to be a standard view of representation. He assumes it is and asks how it can be justified: "By what argument can it be proved, that the perceptions of the mind must be caused by external objects, entirely different from them, though resembling them (if that be possible)... (EHU 152–153). Hume's parenthetical "(if that be possible)" acknowledges the Berkeleyan concern that "an idea can be like nothing but an idea." This concern has no force when perceptions of perceptions are at issue (Berkeley, *Principles*, Part I, Section 8).

41 Presumably Descartes would give a similar explanation with this theory of representation of how objects with only primary qualities, are misrepresented as having secondary qualities (See Meditation III, 158, Meditation VI, 187, *Principles*, Part I, Principle LXXI, 249, *The Passions of the Soul*, Part First, Article xxiii). This misrepresentation may be where Hume got the idea that the mind so readily spreads itself on the world.

42 The context shows that the "in effect" in this quotation could just as well have been left out.

43 I am indebted here to Don Garrett.

44 The foregoing was motivated by an objection of Elise Springer's.

45 For a more detailed formulation of the problem see my "Hume's Puzzle about Identity," forthcoming in *Philosophical Studies*.

46 This problem, *mutatis mutandis*, will apply to representing a recognized one as many also. It is a problem for his general account of the idea of identity.

47 See T 69: "Two objects, tho' perfectly resembling each other, and even appearing in the same place at different times, may be numerically different."

48 Mark Rubin made me aware of this.

49 For example Fogelin (100) lends support to the first, and (105) to the second, and Stroud (128) to the third.

50 I am indebted here also to Don Garrett.

51 There is little explanation in the *Treatise* for why we think of the self as active. Likely the explanation is, as at T 222, that we tend to think of the self as the cause of its perceptions. It would be interesting to see if this could be carried out in detail.

52 For example, he says of the vulgar's attribution of "a distinct continu'd existence to the very things they feel or see" that "this sentiment, then, as it is entirely unreasonable, must proceed from some other faculty than the understanding" (T 193). He is assuming that the explanation for beliefs issuing from the understanding would be that they are reasonable. The "other faculty" is the imagination. The explanation for beliefs issuing from the imagination is, rather, causal.

53 Likewise, he can no longer give a naturalistic explanation of the "fictions of the antient philosophy" (T 219).