RETHINKING HUME’S SECOND THOUGHTS ABOUT PERSONAL IDENTITY

DON GARRETT
NEW YORK UNIVERSITY

In the Appendix to A Treatise of Human Nature, David Hume wrote:

I had entertain’d some hopes, that however deficient our theory of the intellectual world might be, it wou’d 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. (THN App.10)¹

Why did Hume become so dissatisfied with the “former opinions” expressed in his discussion of personal identity? This is an important interpretive question for at least two reasons. First, we cannot claim to understand Hume’s philosophy fully unless we know what problem he thought he saw with his original account of personal identity. Second, when a philosopher of Hume’s penetration claims to find a problem of such seriousness in a central element of his philosophy, there is very likely to be a serious problem that we would benefit from understanding.

¹ Hume 2007. This work is cited as “THN,” with numerical references following to book, part, section, and paragraph, respectively. References to its “Appendix” are indicated as “THN App.” followed by the paragraph number. Italics in quoted passages are in the original. Because the first edition of A Treatise of Human Nature was published in two volumes, there was a significant time gap between the appearance of the “section concerning personal identity,” in Book 1 of the Treatise, and the appearance of the Appendix: the first volume, containing Books 1 and 2, appeared in 1739, whereas the second volume, containing Book 3 and the Appendix, appeared in 1740.
The question is also a difficult one, for Hume’s own characterization of the cause of his dissatisfaction is in key respects frustratingly unspecific. In consequence, it has received what is surely a far greater number of distinct answers—well over two dozen, even by a conservative count—than has any other interpretive question about Hume’s philosophical writings. Fortunately, the answers offered do tend to fall within a few very general classes, distinguished by the kind of problem they interpret Hume as recognizing. Barry Stroud provided one of the best-known answers in his landmark 1977 book, *Hume*. I have offered another (Garrett 1981 and 1997). Because both answers interpret Hume as worried about explaining the metaphysics of the “ownership” of perceptions by individual minds, they are sometimes—quite properly, I think—classified together as instances of the same general type. Indeed, mine is in some ways just a further specification of his.

To my knowledge, no other commentator has ever simply endorsed either Stroud’s answer or mine. Indeed, I believe it is fair to say that no commentator has ever simply endorsed the answer of any other commentator. Despite that rather daunting fact, I propose to enter the fray once more. First, I will outline the structure and content of the relevant primary texts and derive from them a set of criteria by which to judge proposed answers to the question. Second, I will distinguish and assess several types of approach to the question, defend the general approach that Stroud and I both favor against a recent objection, and explain why I find Stroud’s specific answer promising but in need of further specification. Finally, I will formulate three doctrines that are central to and highly salient in “Of Personal Identity,” describe some of the multiple unacceptable consequences their conjunction produces, and argue that this problem alone meets the criteria for a satisfactory answer to the question. In doing so, I hope to vindicate Stroud’s initial insight and
convince at least some readers that the problem I describe is the problem Hume actually saw. As before, I remain quite certain that it is the most important problem that he could have seen—and the most important for us to see.

I. Analysis of Texts and Criteria for Success

“Of personal identity.” The “section concerning personal identity” to which Hume’s Appendix refers—THN 1.4.6, entitled “Of personal identity”—addresses two related questions:

*The Metaphysical Question:* What is the mind, self, or person to which identity and simplicity are attributed?

*The Psychological Question:* How and why are identity and simplicity attributed to a mind, self, or person?

It is important to note two things about the terminology of these questions. First, ‘identity’, for Hume, designates a purely diachronic property that is the absence of both qualitative variation and interruption, while “simplicity” designates a purely synchronic property that is the absence of composition by parts. Second, he is here using the terms ‘mind’ ‘self’, and ‘person’ in narrow senses designating a potentially introspectible mental entity to which identity and simplicity are naturally attributed, even though he is also committed to the existence of what we might call “background machinery,” presumably including the brain and its processes, causally supporting mental operations.²

² Hume is committed to the existence of such background machinery by his universal determinism, since the sequence of thoughts in a mind cannot be predicted by deterministic mental laws alone. As Galen Strawson has rightly emphasized in correspondence, Hume’s doctrine that causes must immediately precede their effects likewise commits him to background machinery behind the operations of memory and habit. See also his discussion (in Strawson 2001) of the need for background machinery to explain the operation of the principles of association. It is a debatable question whether Hume ever uses the term ‘mind’ to designate this background machinery was well. It is also debatable whether he sometimes employs or allows a broader
“Of personal identity” may be conveniently divided into three main subsections.

Subsection 1, consisting of THN 1.4.6.1-4, argues that a mind, person, or self is not something genuinely simple and identical on which perceptions depend, but is instead “nothing but a bundle or collection of perceptions,” lacking the “perfect identity” that is possessed only by things that are “invariable and uninterrupted,” and also lacking the “perfect simplicity” that is possessed only by things that have no simultaneously co-existing parts. There is, Hume claims, no impression constant and invariable throughout the course of one’s life from which the idea of such a simple and identical self could be derived. Rather, one finds only diverse perceptions in “a perpetual flux and movement,” and these perceptions are “different, and distinguishable, and separable from each other, and may be separately consider’d, and may exist separately, and have no need of any thing to support their existence.”

Subsection 2, consisting of THN 1.4.6.5-16, argues that the mind has, instead of perfect identity, only what he calls a “fictitious” or “imperfect” identity, one that results from principles of association that operate in the imagination on relations holding among these perceptions. In explanation of this claim, Hume invokes his general account (already presented in THN 1.4.2, “Of skepticism with regard to the senses”) of how related successions come to constitute fictitious or imperfect identities: they do so through being naturally and almost irresistibly conflated with genuine or perfect identities. This conflation occurs, he asserts, because the feeling of considering (in memory) a succession of related objects is very similar to the feeling of considering an invariable and uninterrupted object. The mind often seeks to “disguise” the error of this conflation, he

sense of ‘self’ that includes the entire body as well as the mind. See particularly THN 2.1.8.1 and 2.1.9.1, and “Dissertation of the Passions” 2.8.
emphasizes, by feigning a supposedly unchanging substantial substratum of what he calls “inhesion” (we would say “inheritance”) to unify the various distinct items of the succession. In the case of personal identity, this supposed substratum would be a substantial soul. In fact, however, the supposed relation of inhesion is unintelligible and the fiction of a substantial soul in which perceptions inhere is, as he has already argued at length in the immediately previous section (THN 1.4.5, “Of the immateriality of the soul”) an “absurdity.” In support of his claim that association-generating relations are responsible for the ascription of identity to the mind, he elaborates on several features of related successions that are conducive to this imaginative conflation, most of which can be readily seen to apply to the succession of ideas that constitutes the mind. He concludes from this elaboration that, despite the rapid succession of its perceptions, the “identity” of a person or mind is fundamentally like that of a plant or animal: that is, it is only an imperfect identity dependent on associative principles of the imagination. The operation of these principles allows the mind to review with a feeling of “ease” or “facility” the succession of its remembered perceptions when these perceptions are (recognized to be) related by one or more of the three relations that he has identified earlier in the Treatise as leading to association: resemblance, contiguity, and causation.

3 Given the characterization of the mind itself as a bundle of perceptions, of course, saying that “the mind” does conflate and feign in this way amounts to saying that such conflation and feigning is, at least in part, a feature of some perceptions in the bundle, their relations, and their causal activity.

4 First, identity is more likely to be ascribed when only a proportionally small part of a complex thing is replaced, or where change occurs slowly and gradually. Second, identity is more likely to be ascribed when the parts of a thing all have reference to a “common end,” and especially where there is a “sympathy of parts” whereby the parts interact with one another, as well as external things, to produce that common end—as they do, for example, in the case of plants and animals. Third, identity can readily be ascribed even when there is a swift and complete replacement of parts, so long as the earlier version of a thing is destroyed before the second is created (as with a church destroyed and rebuilt of different parts) or the thing is naturally expected to undergo ongoing replacement of specific parts (as in a flowing river).
Hume then supports this result by noting that he has already established that the understanding—a potential competitor to “the imagination” and its principles of association\(^5\)—can discern no “real connexions among objects, and that even the union of cause and effect, when strictly examin’d, resolves itself into a customary association of ideas” (THN 1.4.6.16). Each half of this remark refers back to his primary discussion of causation in THN 1.3.14, “Of the idea of necessary connexion.” The denial of “real connexions” discernible by the understanding refers most directly to the following passage, which specifies a test for such connections:

Now nothing is more evident than that the human mind cannot form such an idea of two objects, as to conceive any connexion betwixt them, or comprehend distinctly that power or efficacy by which they are united. Such a connexion would amount to a demonstration, and would imply the absolute impossibility for the one object not to follow, or to be conceived not to follow upon the other: which kind of connexion has already been rejected in all cases. (THN 1.3.14.13)\(^6\)

The claim that the “union” of causes and effects resolves itself into a “customary association of ideas” refers to his argument (THN 1.3.14.16-22) that the idea of a “necessary connexion” between cause and effect is copied from the impression or feeling of a psychological “determination of the mind” to pass from the perception of one to the idea of the other, a determination that itself results from their having become habitually associated in the mind of the observer. Hume allows, of course, that the understanding can discern many relations of various kinds among things. But because it does not perceive a connecting principle between any two distinct objects—even between causes and effects—principles of the imagination that are dependent on association-generating relations remain

---

\(^5\) See Garrett 1997, chapter 1, for a full account of Hume’s distinction between the imagination and the understanding.

\(^6\) The last line of this passage, in turn, clearly alludes to THN 1.3.3, “Why a cause is always necessary.”
the only viable way to account for the mind’s tendency to regard a bundle of related successive perceptions as an instance of identity.

Subsection 3, consisting of THN 1.4.6.17-22, first argues that the specific association-generating relations required for the ascription of identity to minds or persons are resemblance and causation; it next explains why those two relations are prevalent within minds; and it closes by drawing some further consequences. Hume begins by stating that, while contiguity is one of three association-generating relations, it “has little or no influence in the present case,” leaving resemblance and causation as the remaining possibilities. Resemblance, he claims, is able to contribute to the imperfect identity primarily because the mind includes many memories of perceptions that resemble the perceptions themselves, thereby introducing a considerable degree of resemblance into the series of perceptions reviewed. Causation, he continues, is able to contribute because “the true idea of the human mind is to consider it as a system of different perceptions or different existences, which are link’d together by the relation of cause and effect, and mutually produce, destroy, influence, and modify each other”; and after comparing the mind to a “republic,” he reaffirms that “whatever changes [a person] endures, his several parts are still connected by the relation of causation.” Hence, there are many noticeable causal relations within the mind that can contribute to the ascription of identity. Indeed, once the causal relations among remembered perceptions are noted, it can be inferred that

---

7 The portion of the paragraph between this opening and this closing reads:
Our impressions give rise to their correspondent ideas; and these ideas, in their turn, produce other impressions. One thought chases another, and draws after it a third, by which it is expelled in its turn. In this respect, I cannot compare the soul more properly to any thing than to a republic or commonwealth, in which the several members are united by the reciprocal ties of government and subordination, and give rise to other persons who propagate the same republic in the incessant changes of its parts. And as the same individual republic may not only change its members, but also its laws and constitutions; in like manner the same person may vary his character and disposition, as well as his impressions and ideas, without losing his identity. (THN 1.4.6.19)
there have been other perceptions, also causally related to these perceptions, that are themselves no longer remembered but are nevertheless equally perceptions in the same mind. After adding that the susceptibility of mental association to degrees renders all “nice and subtile questions concerning personal identity” merely verbal or “grammatical,” rather than philosophical, Hume concludes by returning to the question of the mind’s simplicity at a time, claiming that an account parallel to that for identity, in terms of associative relations in the imagination, also explains the fictitious or imperfect simplicity ascribed to the complex mind.

Hume’s answers to his two main questions in the section “Of personal identity,” then, are as follows:

*Answer to the Metaphysical Question:* The mind is in fact a bundle of perceptions (Subsection 1), and not a substantial soul (Subsection 2); some of these perceptions resemble others via memory, but what links all of the perceptions in the bundle or “system” into a single mind is the relation of cause and effect (Subsection 3). It is for this reason that unremembered perceptions belonging to this causal system are still properly parts of the mind (Subsection 3).

*Answer to the Psychological Question:* Although the mind lacks perfect identity (Subsection 1), identity is nevertheless attributed to it because relations among its elements generate mental association and so make the feeling of reviewing those elements in memory similar to the feeling of reviewing the perfect identity of an unchanging and uninterrupted object (Subsection 2). This constitutes a “fictitious” or “imperfect” identity for the mind (Subsection 2). The specific relations leading the imagination to ascribe identity to the mind are resemblance and causation (Subsection 3). A similar account explains the ascription of simplicity to the mind (Subsection 3).
Thus, the mind is a system of causally-related complex successive perceptions; as such, it would remain what it is regardless of whether identity or simplicity were ever ascribed to it. In fact, however, it naturally provokes erroneous ascriptions of perfect identity and simplicity, and it can thereby be said to have a kind of imperfect or fictitious identity and simplicity.

The Appendix. Of the twenty-two paragraphs in the Appendix to the Treatise that are additions marked for insertion into the main text, more than half, THN App.10-21, are devoted to personal identity and simplicity. In the first of those paragraphs, Hume announces, in language already quoted, his discovery of a “contradiction” or inconsistency in the course of reviewing the section “Of personal identity,” and he then promises to “propose the arguments on both sides, beginning with those that induc’d me to deny the strict and proper identity and simplicity of a self or thinking being.” The next nine paragraphs, THN App.11-19, offer a lengthy series of short arguments (some old, some new) for this denial, thereby reinforcing the conclusion of Subsection 1 of “Of personal identity.” It is left to the next paragraph, THN App.20, to present the promised other “side.” That paragraph, however, contains no argument that the mind does have strict and proper (i.e., perfect) identity and simplicity. Instead, Hume simply repeats his argument from THN 1.4.6.16—that is, from the final paragraph of Subsection 2—that the ascription of identity must result from principles of association rather than from a discovery of any “real connexion” among perceptions. He does so, moreover, in between two similar general descriptions of when and how the new problem that he now sees has arisen. The full paragraph is as follows (note the placement of Hume’s own numbered footnote superscript, to be discussed in the next section):
So far I seem to be attended with sufficient evidence. But having thus loosen’d all our particular perceptions, when I proceed to explain the principle of connexion, \(^1\) 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. If perceptions are distinct existences, they form a whole only by being connected together. But no connexions among distinct existences are ever discoverable by human understanding. We only feel a connexion or a determination of the thought, to pass from one object to another. It follows, therefore, that the thought alone finds personal identity, when reflecting on the train of past perceptions, that compose a mind, the ideas of them are felt to be connected together, and naturally introduce each other. However extraordinary this conclusion may seem, it need not surprize us. Most philosophers seem inclin’d to think that personal identity arises from consciousness; and consciousness is nothing but a reflected thought or perception. But 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. (THH App.20)

The final paragraph devoted to personal identity and simplicity then sums up and re-describes the difficulty in terms of two principles:

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. Did our perceptions either 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. For my part, I must plead the privilege of a sceptic, and confess that this difficulty is too hard for my understanding. I pretend not, however, to pronounce it absolutely insuperable. Others, perhaps, or myself, upon more mature reflection, may discover some hypothesis, that will reconcile these contradictions. (THN App.21)
The first of these two “Unrenounceable Principles,” as we may call them, clearly derives from Subsection 1 (supplemented by the explicit rejection of substantial souls in Subsection 2) and the long first “side” of the Appendix; the second principle clearly derives from the final argument of Subsection 2 and the short second “side” of the Appendix.

The two Unrenounceable Principles are not strictly inconsistent with each other, of course; nor, as Fogelin (1992) points out, could they possibly be “reconciled” by any new “hypothesis” if they were. Rather, Hume quite evidently sees only three jointly exhaustive alternatives for explaining whatever kind of identity and simplicity—either perfect or imperfect—the mind has: (i) the mind’s perceptions are not distinct existences but instead inhere in a mental substance; (ii) the mind’s perceptions are related by a “real connexion” that is perceived by the understanding; or (iii) the mind’s perceptions stand in association-generating relations that give rise to a felt unity in the imagination when they are reviewed in memory. The first alternative would make the mind a substance with perfect identity and simplicity; the second and third would allow some kind of imperfect identity and simplicity to a bundle of perceptions. A standoff now seems to arise because, while still convinced that the first two alternatives cannot be right, he has suddenly lost confidence in the third and only remaining alternative. To render the inconsistency more explicit, therefore, we must add to the two Unrenounceable Principles the following three supporting claims:

---

8 It may seem that there is a fourth alternative: that the mind’s perceptions are related by a “real connexion” that is not perceived by the understanding. But whereas inhesion is a soul would account for the mind’s identity and simplicity whether perceived or not, only a perceived relation could explain the mind’s tendency to attribute identity and simplicity to a succession of perceptions that in fact lacks them.
(A) The identity and simplicity of the mind, whether perfect or imperfect, results either (i) from its being something (such as a substantial soul) that prevents distinct perceptions from being distinct existences or (ii) from a perceived relation among perceptions as distinct existences.

(B) Any relation among distinct perceptions from which the identity and simplicity of the mind could result is either a “real connexion” perceived by the understanding or else an association-generating relation that produces a felt unity of the perceptions in the imagination.

(C) The identity and simplicity of the mind does not result from association-generating relations that produce a felt unity of the perceptions in the imagination.

Hume treats claims (A) and (B) as entirely settled in the section “Of personal identity,” and they remain settled throughout the Appendix, at least up until the expressed hope for some new hypothesis in the final sentence of THN App.21. While (C) runs directly contrary to “Of personal identity,” it is now strongly implied by Hume’s admission in THN App.20 that his “account is very defective” and that he lacks any theory that will give him “satisfaction on this head.” At the very least, he implies that the identity and simplicity of the mind cannot be understood to result from such relations, whereas he had thought that it could. But why, specifically, has Hume concluded that his previous theory was defective and unsatisfactory? He does not actually say—hence the puzzle.

Criteria for a Successful Answer. Although Hume does not explicitly answer the question of why he finds his previous theory unsatisfactory, the structure and content of his remarks in the Appendix suggest five criteria that an account of the problem should meet:

Crisis Criterion: The problem should fit Hume’s initial description of it in the Appendix—that is, it should concern a problem that, had he noticed it, would
have appeared to him to be a very serious one, definitively threatening contradiction and absurdity that would involve him in a “labyrinth” and be sufficient to inspire skepticism.

*Origin Criterion:* The problem should fit his description of how it arises—that is, it should naturally arise or be recognized, at least upon review, when proceeding “to explain the principle of connexion, which binds [the mind’s perceptions] together, and makes us attribute to them a real simplicity and identity” and coming “to explain the principles, that unite our successive perceptions in our thought or consciousness.”

*Solution Criterion:* The problem should fit his concluding description of it, as involving the two Unrenounceable Principles and as being at least seemingly soluble if “perceptions inhere in something simple and individual” or “the mind perceived some real connexion among distinct existences.”

*Scope Criterion:* The problem should be one that Hume could plausibly have seen as applying primarily to his account of personal identity and simplicity, and not (for example) equally to his account of all ascriptions of identity and simplicity, or to his account of mental life more generally—for he expresses worry only about his account of personal identity and simplicity.

*Difficulty Criterion:* It would be at least desirable if the problem were one that, while within Hume’s ability to notice upon review, would be at least somewhat difficult for him to express clearly—for despite his enormous gifts as a writer, he does not manage to say what it is.

### II. Types of Answer Compared

*Four Types of Answers.* Stroud distinguishes two ways to explain the source of Hume’s second thoughts, based on two ways to interpret the Appendix description of how the new-found problem arises:
His ‘hopes vanish’ only when he comes to ‘explain the principles, that unite our successive perceptions in our thought or consciousness’. This statement of the difficulty is ambiguous. It could mean that Hume has no hope of explaining what actually unites our successive perceptions into one mind or consciousness—what actually ties them together to make up one mind. Or it could mean that he has no hope of explaining what features of our perceptions and what principles of the mind combine to produce in us the thought or belief that we are individual minds—what ties the successive perceptions together in our thought, or what make us think of them as tied together. Obviously, these two interpretations are different. (133; emphasis in original)

Let us use the term “metaphysics-of-bundling answers” to designate the type of answer that interprets Hume’s problem in the first way—that is, as concerned at least in part with how perceptions are actually linked together to make up a mind in which or on which associative mechanisms can operate. In contrast, answers that interpret Hume’s problem in the second way, as concerned solely with the operations of the psychological principles that explain our ascriptions of identity and simplicity to a mind, may be called “psychology-of-ascription answers.” Among psychology-of-ascription answers, we may further distinguish two types: those that locate Hume’s problem in the operations of general psychological principles other than the associative principles of resemblance and causation, and those that locate the problem specifically in something about the scope or operation of those two associative principles. Finally, we may distinguish a miscellaneous type of answer that locates Hume’s problem in something else entirely, neither in the metaphysics of bundling nor in the psychological mechanisms of identity ascription. These four types of answers correspond, at least roughly, to what Jonathan Ellis (2006) has recently dubbed “Group 1,” “Group 2,” “Group 3,” and “Group 4” interpretations, respectively.
Particular answers, of any type, to the question of the source of Hume’s second thoughts may in principle run afoul of any of the criteria for a successful answer; many run afoul of more than one. Given the number of different answers in print, it has become impractical to consider them all individually. (Garrett 1997 examines and argues against eight; Garrett 2009b argues against a ninth.) Ellis’s typology, however, allows him—and us—to consider objections to multiple answers at once. Taking the last type (“Group 4”) first, answers that concern neither the metaphysics of bundling nor the psychology of identity ascription can, he notes, generally give 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”; hence, they will fail to satisfy at least the Origin Criterion. They will also typically fare poorly by the Solution Criterion.

Psychology-of-ascription answers that interpret Hume’s problem as concerned with the operations of psychological principles other than the associative principles of resemblance and causation (“Group 2”) will likewise not fare well by the Origin Criterion, since Hume seemingly does not “come to explain” any such further principles after having “loosen’d all our particular perceptions” in the section “Of personal identity.” Such answers often run afoul of the Scope Criterion as well, since the problems alleged with the psychological principles in question generally implicate more than just Hume’s account of personal identity and simplicity. Psychology-of-ascription answers that do pertain specifically to resemblance and causation as association-generating relations (“Group 3”), in contrast, can often do well by the Origin Criterion, but historically have failed to satisfy the Crisis

---

9 Ellis mentions Kemp Smith 1941, Penelhum 1951, and Swain 2006 as examples of Group 4 interpretations.
Criterion: in the past, none has described a problem that Hume would have thought to pose a serious problem or “contradiction” had he seen it. Ellis helpfully supports this general verdict about answers of this type by considering six different instances of it (four of them fairly recent) and showing in each case why Hume would not have been worried about the problem it describes.\textsuperscript{11}

Ellis himself proposes a new answer that he intends to be of this type, one that he believes does describe a problem that Hume would find insurmountable. This problem concerns the origin of what Ellis calls the “fictitious idea of the self.” According to Ellis, Hume realizes that the associative principles of resemblance and causation are insufficient to generate the idea of the self because he has already denied that we ever perceive—that is, have an impression of—the self. Since ideas must be copied from impressions for Hume, no matter how readily the mind passes from one idea to another or how thoroughly their objects become associated, no reflective review process could ever give rise to the fictitious idea of a self.

Ellis ultimately concedes, in effect, that this answer itself runs into difficulty with the Origin Criterion, since the problem described is largely with explaining the content of the idea of the self in light of Hume’s requirement that all ideas be derived from impressions, rather than with the associative principles of resemblance and causation as such. More deeply, however, the answer appears to misconstrue both Hume’s psychological task and his psychological resources. Contrary to Ellis’s suggestion, Hume never characterizes the idea of the mind or self—as opposed to the supposed idea of the soul as a substratum of

\textsuperscript{11} These six interpretations are those of Patten 1976, Haugeland 1998 (which was well known for many years before its formal publication), Baxter 1998, Roth 2000, Winkler 2000, and Ainslie 2001. I would also classify Baier 2008 (Chapter 9), published after Ellis’s article, in this group. Baier locates the difficulty primarily in applying Hume’s second definition of ‘cause’ to perceptions themselves, given that there are not “impressions” of perceptions. I think that this problem, too, does not meet the Crisis Criterion.
inhesion—as “fictitious.” On the contrary, as we have seen, he says explicitly what the
“true idea of the human mind” is: it is the idea of “a system of different perceptions or
different existences, which are link’d together by the relation of cause and effect, and
mutually produce, destroy, influence, and modify each other.”12 It is not the idea of the
mind or self that he calls “fictitious,” but only the mind’s simplicity at one time and identity
through time. (Compare: the simplicity and identity of a complex and changing tree are
also fictitious, for Hume, but the true idea of the complex and changing tree understood as
a composite succession is not itself a fiction.) 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
states explicitly that “the idea, or rather impression, of ourselves is always intimately
present with us” (THN 2.1.11.4).13 It is only an impression of the self as something simple
and identical or as distinct from our perceptions that he claims not to find by introspection.
Any perception in the mind, as part of the bundle of perceptions that constitutes a mind, is
a perception “of” the mind. (Note that a perception, for Hume, may be “of” more than one
thing, such as a color and a shape—or a passion and the self.) Once again, we do not have
a problem that would have worried Hume.

Metaphysics-of-Bundling Answers. This brings us back to metaphysics-of-bundling
answers. Ellis rejects all answers of this type on the grounds that they, too, fail to satisfy

---

12 It is true that the mind cannot retain ideas of all of these perceptions at a single time, but Hume’s theory of
abstract ideas (THN 1.1.7; see Garrett 1997, Chapter 8) explains how one or more exemplar ideas can
represent all of a set of things that have something in common—including, in this, the perceptions that
constitute a causally-connected system or bundle.

13 The theory of abstract ideas also explains how any idea of one of our perceptions can function as the
(exemplar) idea of self, and hence how any impression in one’s own mind can serve as an impression of the
self (that is, as one of the impressions from which the idea of the self can be derived; see Garrett 1997).
Ellis’s answer has strong affinities with that of Kemp Smith (1941), who argued that Hume’s contradiction
consisted of both affirming and denying that there is an impression of the self.
the Origin Criterion. This, he argues, is because close attention to the content of THN App.20 shows that Hume’s concern is not with the metaphysical question of how perceptions are actually related so as to constitute bundles. Ellis’s argument may be reconstructed as follows:

(1) In the tenth sentence of THN App.20, when Hume describes all his hopes vanishing when he “comes to explain the principles, that unite our successive perceptions in our thought or consciousness,” he must be understood as referring to the same principle or principles to which the second sentence of THN App.20 refers.

(2) In the second sentence of THN App.20, Hume refers solely to the operations of the psychological principle of association, and not to a metaphysical principle of unification, when he writes, “But having thus loosen’d all our particular perceptions, when I proceed to explain the principle of connexion,¹ which binds them together, and makes us attribute to them a real simplicity and identity; I am sensible that my account is very defective.” [Note: I follow Norton and Norton (Hume 2007) in placing the footnote after ‘connexion’; Ellis, following the earlier Selby-Bigge/Nidditch edition, places it after ‘when’. I do not see that this makes any difference to the interpretation.]

(3) Hence, the cause of Hume’s “vanishing hopes” lies entirely in 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. [from (1) and (2)]

Ellis’s argument for (1) appeals quite convincingly to various textual details of “Of personal identity” and the Appendix. His argument for (2) appeals to the footnote that Hume inserted following the word ‘when’ in the second sentence of THN App.20. This footnote, as it appears in the first edition of the Treatise (the only edition published during

---
¹ Ellis classifies Basson 1958, Beauchamp 1979, Pears 1975, Stroud 1977, and Garrett 1981 (and 1997) as Group 1 interpretations. I would add Strawson 2001 and, more recently, Kail 2007 to this group.
Hume’s lifetime), refers to “page 452.” That page of the first edition contains the final lines of THN 1.4.6.16, all of THN 1.4.6.17, and the beginning of THN 1.4.6.18; in other words, it marks the transition from Subsection 2 of “Of personal identity” to Subsection 3, in which Hume proceeds to explain the principles that unite our perceptions in the imagination and thereby produce an ascription of identity.

Clarifying the precise reference of Hume’s footnote is very useful; it is not, however, sufficient to establish (2). Ellis writes that “Hume’s primary task in THN 1.4.6 is to explain how we form the idea of a self on the basis of our impressions and those of our ideas the formation of which Hume has already explained” (208), and he suggests that the question of what actually links the perceptions into the bundle constituting the mind does not arise. As we have seen, however, Hume has two primary explanatory aims in the section: to explain what the human mind, person, or self is, and how identity and simplicity are ascribed to it. (Were he concerned only with the origin of the idea of the self, he could have called the section “Of the idea of self;” just as he calls THN 1.2 “Of the ideas of space and time,” THN 1.3.2 “Of probability; and of the idea of cause and effect,” and THN 1.3.14 “Of the idea of necessary connexion.” In fact, none of the sections of THN 1.4 [“Of the skeptical and other systems of philosophy”] is directed exclusively at explaining the origin of an idea.) Meeting the first explanatory aim is a crucial element in meeting the second. Thus, Subsection 3—the line of argument to which Hume “proceeds” on the page to which his footnote refers us—argues that resemblance and causation are the association-generating relations responsible for the ascription of identity and simplicity to the mind; but it also goes on immediately to explain why these relations are able to play this role. Specifically, it does so by appealing to the constitution of the human mind itself as a
bundle of causally-related perceptions that includes memories. Ellis assumes that
metaphysics-of-bundling answers must interpret Hume’s reference to a “principle” or
“principles” in the second and tenth sentences of THN App.20 as referring exclusively to
metaphysical principles of bundling rather than to psychological principles of association;
but that is not the case. In fact, Hume may well be thinking in the second sentence of
causation as a “principle” that “binds” perceptions “together” metaphysically, and “makes
us attribute to them a real simplicity and identity” psychologically. But even if he means
by the terms ‘principle’ and ‘principles’ just the psychological principles accounting for
ascriptions of identity,¹⁵ his attempted “explanation” of these principles—which he
identifies as the source of his despair—crucially invokes the actual causal relations among
perceptions that link them into bundles. In reviewing this subsection, therefore, Hume may
well have realized that he was unable in some way to explain this bundling satisfactorily. If
the metaphysics-of-bundling approach is correct, that is exactly what happened.

Stroud’s Answer. Stroud characterizes the problem that he thinks most likely to be the
cause of Hume’s worries as “elusive”:

(i) [It is] an undeniable fact that the only ‘data’ available to a person for the
formation of his ideas and beliefs are his own perceptions…. But although
the fact seems undeniable, Hume has no way of accounting for it…. He
cannot explain how or why the ‘data’ from which the idea of personal
identity is constructed present themselves in the way they do. And if they
did not present themselves that way, his explanation would collapse. The
point is elusive, and I think I can only indicate roughly what I have in mind.
(136-37)

¹⁵ Strawson 2001 rightly observes that explaining the psychological principles leading to ascriptions of
identity is likely to involve more than just identifying and describing those principles.
For the sake of maximum accuracy, therefore, I will quote Stroud’s subsequent elaborations at some length:

(ii) It is therefore clear that Hume’s explanation of the origin of the idea of the self or mind is not necessarily deficient in failing to give an account of how a certain idea raises from certain ‘data’, but that it leaves completely unintelligible and mysterious the fact that those ‘data’ are as they are. When we press on to that level of inquiry we find it simply taken as a given fact about the universe of perceptions that the range of reflective vision of any one of them does not extend to all the rest. And it is only because one’s gaze is thus restricted to a certain subset of all the perceptions there are that it is possible for a person to get an idea of himself.

But why is our gaze restricted in that way? What accounts for the fact that one cannot survey in the same way all the perceptions there are? It seems as if one’s vision could be non-circularly restricted to a certain subset or series only if perceptions in fact occurred only as members of particular series—as if they came already tied together into bundles, as it were, so that no member of a particular series could be a perception of any perception outside that series. But how could that be, on Hume’s theory? For him all perceptions are distinct existences, and so each one could exist independently of every other and independently of everything else in the universe. 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. (139)

(iii) 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 simple and individual, or did the mind perceive some real connexion among them, there wou’d be no difficulty in the case.’… If there were such things as
simple spiritual substances to which perceptions necessarily belonged, or if there were ‘real’ or necessary connections between certain perceptions, then that would perhaps explain why the field of vision of a particular reflective perception is restricted to a certain subset of all the perceptions there are…. In short, it would explain why the ‘data’ are in fact the way they must be if anything like Hume’s explanation of the origin of the idea of the self is to be successful. (139)

(iv) [T]he scope of one’s experience does not extend to all the perceptions there are, and that is the inexplicable fact upon which Hume’s explanation depends. To say it is ‘inexplicable’ for Hume is to say that it is inconsistent with the theory of ideas [specifically, as it entails the two Unrenounceable Principles], which he takes to be the only way to make sense of psychological phenomena. (140)

As Stroud emphasizes, in order for the associative principles to operate properly in helping to produce the idea of the self, perceptions must belong to bundles such that earlier members are reflectively accessible (that is, can be remembered) only within the same bundle; the worry he attributes to Hume is that this fact is “inexplicable,” given that there are other perceptions in the universe. Substantial souls or necessary causal “real connexions” might well serve to explain this fact, if there could be such things. But while Hume does not recognize perceptible “real connexions” rendering causes and effects “demonstrably” or “metaphysically” inseparable, he does recognize genuine causal relations; the Treatise, as an investigation into the “science of man,” is an attempt to discover many of them. So why should the “inexplicable fact” about memory access not be explained simply by pointing out that perceptions come quite generally in causally discrete bundles—that is, bundles such that only members of the same bundle have effects,
unmediated by external bodily motions, on other perceptions in the same bundle? After all, Hume claims that the human mind is a “system of different perceptions … link’d together by the relation of cause and effect.” It may well be that causation is a relation holding among metaphysically distinct entities, for Hume, and not an “intrinsic” feature of them (as Stroud notes in passage [ii] above); but why should this matter, as long as the relations do actually exist and tie perceptions into bundles of just the kind Stroud indicates are needed to allow appropriate reflective access?

Of course, it might be argued that Hume would still need a further explanation of why there are such causally discrete bundles. But what kind of explanation would this be? If the explanation is itself causal, then our lack of knowledge of it could hardly be something that Hume would consider to be a crisis or contradiction, for he freely acknowledges much ignorance of causes. Furthermore, an obvious explanatory causal hypothesis does present itself: many or all perceptions seemingly have causes and effects in a particular brain—a different brain for each person. For all that we have seen so far, perceptions may come in causally discrete bundles because they are causally dependent on separate brains. If Stroud is on the right track, then it seems that there must be some as yet unspecified further problem with the causal relations that, according to Hume’s original account, are responsible for tying perceptions into bundles. I will now try again to specify what that problem is.

III. Mental Places and Mental Causes

16 I distinguish external bodily motions—which are sensible by other minds—from internal bodily motions such as events in the brain or central nervous system. Jonathan Cottrell helped me to see the importance of this distinction in this particular context.
Immediately after concluding that the mind is “nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement,” Hume offers a telling analogy:

The mind is a kind of theatre, where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations. There is properly no *simplicity* in it at one time, nor identity in different; whatever natural propension we may have to imagine that simplicity and identity. The comparison of the theatre must not mislead us. They are the successive perceptions only, that constitute the mind; nor have we the most distant notion of the place, where these scenes are represented, or of the materials, of which it is compos’d. (THN 1.4.6.4)

What Hume came to realize upon later review, I propose, was simply that his own conception of the mind did after all *require* just what he remarked he could not provide: “a notion of the place” where all and only the perceptions of each mind occur. This realization resulted from his considering in close proximity three of his most central doctrines, all highly salient in “Of personal identity,” that together have a number of consequences that he would have found unacceptable.

*Three Central Doctrines*. The first central doctrine implicated in Hume’s problem is

*Placeless Perceptions*: No non-visual and non-tactile perception is in any “place,” either spiritual (such as a soul or mental substance) or spatial, by which it is located relative to any other perception. Even visual and tactile perceptions are not in any place by which they are located relative to any other perceptions except to those (if any) with which they form a spatially complex perception.

How are some simultaneous perceptions able to constitute the perceptions of a single mind at any given time? And how are some successive perceptions able to constitute the perceptions of a single mind through time? One standard response to these questions is that
perceptions constitute those of a single mind by inhering in a single substantial soul that endures through time; souls thereby provide a kind of spiritual “place” for all and only the perceptions of a single mind. Hume rejects this response—in “Of personal identity” itself, as we have seen, but also in the previous section, THN 1.4.5, “Of the immateriality of the soul,” where he writes:

    We have, therefore, no idea of a substance. Inhesion in something is suppos’d to be requisite to support the existence of our perceptions. Nothing appears requisite to support the existence of a perception. We have, therefore, no idea of inhesion. (THN 1.4.5.6)

Hume also holds, however, that some perceptions—namely, visual and tactile ones—can stand in spatial relations to one another in spatial complexes. It is important to emphasize that these perceptions are not merely of spatially-related things, having spatial arrangement as part of their intentional content without themselves having any literal extension of their own, as Cartesian ideas (and, at least probably, Lockean ideas) do; Humean visual and tactile perceptions quite literally have sizes, shapes, and spatial arrangements in their own right, in just the same sense that bodies do. While Hume is entirely consistent about the literal spatiality of visual and tactile perceptions, his commitment to it is often missed by readers of the Treatise through neglect of the sections in which it does the most work, most notably THN 1.2.1-2 (where ideas are actually compared with bodies in size) and THN 1.4.4-5 (where it is emphasized that some

---

17 Hume is committed to the existence of four kinds of spatial perceptions: visual impressions, tactile impressions, visual ideas, and tactile ideas. He does not explicitly consider whether any complexes of one of these four kinds ever stand in spatial arrangements relative to complexes of another—for example, whether a tactile impression of heat in one’s finger can be quite literally located relative to the visual impression of one’s finger, or whether a visual idea of an elephant can be located somewhere in relation to one’s visual impression of the room in which one is sitting. Fortunately, however, his views on this matter have no effect on the present argument.

18 For an account of how Humean perceptions have representational content, when they do, see Garrett 2006.
perceptions are extended and others are not). In virtue of their arrangement into spatial complexes, there is an obvious sense in which some perceptions in a mind do share a “place” with those others that are parts of the same spatially complex perception. Because of this fact, “contiguity” is able to contribute as an association-generating relation at least to the ascription of simplicity at a time to some perceptions in the mind. Crucially, however, these limited spatial relations are not enough to give a common place to all of the perceptions that exist in a mind at a given time. For as Hume argues in “Of the immateriality of the soul,” the mind also includes many perceptions—impressions and ideas of sounds, tastes, and smells, as well as passions and other sentiments, together with ideas of these—that have no spatial relation to any other perceptions nor to anything else. They are, as he says “no where”:

An object may exist, and yet be no where; and I assert that this is not only possible, but that the greatest part of beings do and must exist after this manner. An object may be said to be no where, when its parts are not so situated with respect to each other, as to form any figure or quantity; nor the whole with respect to other bodies so as to answer to our notions of contiguity or distance. Now this is evidently the case with all our perceptions and objects, except those of the sight and feeling. A moral reflection cannot be placed on the right or on the left hand of a passion; nor can a smell or sound be either of a circular or a square figure. These objects and perceptions, so far from requiring any particular place, are absolutely incompatible with it, and even the imagination cannot attribute it to them….

It will not now be necessary to prove, that those perceptions, which are simple, and exist no where, are incapable of any conjunction in place with matter or body, which is extended and divisible. (THN 1.4.5.10-11; see also THN 1.3.14.25)

As Hume puts the point near the end of the section, “All our perceptions are not susceptible of a local union” (THN 1.4.5.33). When taken together with his rejection,
earlier in the same section, of the soul as a subject of “inhesion,” this implies that there is no sense in which all of the perceptions of a single mind can be understood to share a common “place.” This result would therefore be highly salient in his review of the *Treatise* just at the point at which he comes to reconsider his account of personal identity and simplicity; moreover, it is dramatically reinforced in the analogy of the theater with which he begins “Of personal identity” itself.

The second central doctrine implicated in Hume’s problem is

*Conjunctive Causation:* Taken together, the following are individually necessary and jointly sufficient for the existence of a causal relation between two objects: (i) priority in time; (ii) contiguity in time and, *where applicable*, in place; and (iii) constant similar conjunction of like objects.

While a Humean “definition” of a term need not be synonymous with the term defined, it does provide necessary and sufficient conditions for the term’s proper application (at the very least to the actual world; see Garrett 1997, Chapter 5, and Garrett 2009a). Hence, Conjunctive Causation is entailed by Hume’s famous first definition of ‘cause’:

We may define a *cause* to be “An object precedent and contiguous to another, and where all the objects resembling the former are placed in like relations of precedency and contiguity to those objects, that resemble the latter.” (THN 1.3.14.31)

Even prior to establishing Placeless Perceptions, Hume explicitly warns that spatial (as opposed to temporal) contiguity cannot be required in the case of unlocated things that are “not susceptible to juxta-position and conjunction” and provides a footnote reference forward to “Of the immateriality of the soul” (THN 1.3.2.7 and 1.3.2.7n; see also THN 1.3.14.25 and 1.3.14.25n) by way of explanation.
Hume’s commitment to Conjunctive Causation is particularly evident in three applications he makes of his definition of ‘cause’. The first lies in the first two “corollaries” (of four in all) that he draws from his definitions of ‘cause’. These two corollaries are, first, “that all causes are of the same kind” and, second, that “there is but one kind of [causal] necessity” (THN 1.3.14.32-33). In support of these corollaries, he argues that causation and causal necessity, respectively, are present if and only if the definition of ‘cause’ is satisfied, writing, for example:

For as our idea of efficiency is deriv’d from the constant conjunction of two objects, wherever this is observ’d, the cause is efficient; and where it is not, there can never be a cause of any kind. For the same reason we must reject the distinction betwixt cause and occasion, when suppos’d to signify any thing essentially different from each other. If constant conjunction be imply’d in what we call occasion, ’tis a real cause. (THN 1.3.14.32; SBN 171)

Hume uses a similar argument in THN 2.3.1.4, “Of liberty and necessity,” to show that human actions are causally necessary if and only if the same definition is satisfied, regardless of any supposed “real connexion” or its absence:

Here then are two particulars, which we are to consider as essential to necessity, viz. the constant union and the inference of the mind; and wherever we discover these we must acknowledge a necessity…. [t]he actions of matter have no necessity, but what is deriv’d from these circumstances. (THN 2.3.1.4; SBN 401)

Most importantly, Hume concludes “Of the immateriality of the soul” itself by arguing at length that matter can be the cause of thought precisely because Conjunctive Causation is true. He writes:

[A]ll objects, which are found to be constantly conjoin’d, are upon that account only to be regarded as causes and effects…. And as the constant conjunction of objects constitutes the very essence of cause and effect, matter and motion may
often be regarded as the causes of thought, as far as we have any notion of that relation. (THN 1.4.5.32-33)

Like Placeless Perceptions, then, Conjunctive Causation would thus be highly salient to Hume in his review of his account of personal identity for two reasons. First, he had just devoted extensive attention to it in the immediately preceding section; and second, he had reinforced it in the section “Of personal identity” itself—in this case, by his argument at the end of Subsection 2 of “Of personal identity” that the understanding never perceives any “real” causal “connexion” between objects.¹⁹

The third central doctrine implicated in Hume’s problem is:

*Causal Bundling*: Perceptions are in the same mind if and only they are elements in a system of relevant causal relations holding among them.

Hume’s positive answer to the question of how distinct perceptions constitute a mind is Causal Bundling: the perceptions are all related—both through time and at a single time—by causation, thereby constituting a causal system in which earlier elements contribute to the production of later elements. He states this doctrine explicitly in his description of “the true idea of the human mind” at the beginning of Subsection 3—precisely the point at which he is beginning to “explain” the principles that unite perceptions in our thought or consciousness. That he regards these causal relations as necessary and sufficient for membership in a mind is shown by his explanation of how we attribute unremembered perceptions to our own minds—precisely by attributing membership in the causal system to them. It is natural to assume that direct causal relations among perceptions are

---

¹⁹Ironically, Kail’s metaphysics-of-bundling answer (2007) depends on attributing to Hume a doctrine incompatible with Conjunctive Causation. This doctrine is Causal Realism—the doctrine that every causal relation must involve a real necessary connection between cause and effect, even if it is unknown and broadly unintelligible to us. There are, I believe, several other difficulties with Kail’s proposal in addition to its incompatibility with Hume’s commitment to Conjunctive Causation.
facilitated by the fact that the perceptions are located in the same mind. Causal Bundling requires, however, that the “location” of perceptions in the same mind instead depends on the existence of direct causal relations among them.

Five Unacceptable Consequences. When he originally composed “Of personal identity,” I suggest, Hume tacitly operated with a picture of each mind as a kind of separate “place” in which that mind’s perceptions uniquely share at least their own common “stage” for causal interaction, if not also a separate “audience” or surrounding “theater.” He was encouraged in this, one may suppose, not only by the residual effects of the common tendency to postulate that each mind is a single mental substance but also by the fact that some perceptions in the mind at a time are indeed spatially related to some others at that time. When he considered Causal Bundling directly in the light of the constraints placed on causal relations by Conjunctive Causation and the constraints placed on contiguity of place by Placeless Perceptions, however, he realized that he needed a conception of the place of perceptions in a mind that he could not provide. The need for such a conception and its simultaneous unavailability are demonstrated by at least five consequences that Hume would have found unacceptable.

The first, and perhaps most fundamental, unacceptable consequence is the Likely Non-Existence of Minds. In the material or bodily realm, Conjunctive Causation’s contiguity-of-place requirement rules out the vast majority of what occurs in the universe just before a physical event as irrelevant to its causation. Without such a requirement, it is likely that every potential physical law would be falsified by co-existing objects; for example, the presence of water one thousand miles away from a blazing fire would in that case falsify any causal law to the effect that water extinguishes fire. But Placeless Perceptions imposes
severe restrictions on the applicability of Conjunctive Causation’s contiguity-of-place requirement to perceptions. Let us use the term ‘unit perception’ for any perception that is either (i) non-visual and non-tactile (and hence is “no where”) or (ii) is a maximal spatial complex of visual and/or tactile perceptions. Then we can say that the contiguity-of-place requirement is inapplicable to any causal relations between any unit perceptions. To take an example (suggested by a discussion in Fogelin 1992) that would be particularly important to Hume: multiple successive perceptions of conjunctions of one type of event with another very often occur without generating an associative link between the two types of events—namely, whenever the successive perceptions of conjunctions do not occur in the same mind. Without a contiguity-of-place requirement, therefore, there can be no true causal laws governing such links, any more than there can be true causal laws governing the extinction of fire without such a requirement. Without causal laws, on Hume’s account of causation, there can be no causal relations. But it follows from Causal Bundling that if there are no causal relations among perceptions, there are no minds either.

The second unacceptable consequence is the Tenuous Character of Co-Perception in a Mind. It follows from Causal Bundling and Conjunctive Causation that two simultaneous unit perceptions can belong to the same mind only if they are joint effects of a previous common cause, or joint causes of a subsequent common effect, or both. This is because Causal Bundling requires that all perceptions in the same mind be related causally (directly or indirectly) as part of a single causal system, whereas Conjunctive Causation entails that a cause cannot be simultaneous with its effect.20 Counterintuitively, then, whether two simultaneous perceptions belong to the same mind or not depends of what occurs before

---

20 Hume also argues independently that causes cannot be simultaneous with their effects at THN 1.3.2.7.
them and after them. Even worse, Placeless Perceptions—which of course helped to motivate Causal Bundling in the first place—requires that it depends not simply what happens “in the same place,” but throughout the universe. This applies even to perceptions that Hume supposes to be intimately but non-spatially related to each other in the same mind, such as a desire (which he regards as a non-representational impression) and the representation of its object (which he regards as an idea). When a desire occurs simultaneously with multiple ideas, which idea (if any) is the idea of its object will depend on the existence of suitable “contiguity-independent” causal laws establishing one of the ideas as a common cause or common effect with the desire. Similarly, which sequence of perceptions constitutes the object of the simultaneous feeling of “facility” or “ease” that is said to contribute to ascriptions of identity will depend on joint causes and/or joint effects of the perceptions and the feeling, respectively. Indeed, a sensation of red and a simultaneous abstract idea of government, or a memory of a sensation of red and a simultaneous memory of an abstract idea of government, can belong to the same mind only if they are involved in a common system of prior causes or subsequent effects.

The third unacceptable consequence is the Failure of Unique Ownership. Hume’s three doctrines entail that qualitatively similar and simultaneous unit perceptions cannot exist in different minds. Instead, one of them will belong to a given mind if and only if the other does as well. For Conjunctive Causation and Placeless Perceptions together entail that no two similar and simultaneous unit perceptions can differ in their causal relations, and Causal Bundling entails that no two perceptions can differ in the mind or minds to which they belong without differing in their causal relations. This consequence applies even to simple and extremely common unlocated impressions such as joy, grief, desire, and
aversion; and it also applies to common feelings of ease or facility themselves, which are meant to explain the ascription of identity to the series of remembered perceptions in different minds.

The fourth unacceptable consequence is the *Failure of Unique Bodily Relations*. Hume’s three doctrines entail that two similar and simultaneous unit perceptions cannot be causally related to different brains or bodies. Instead, one of them will be causally related to a given brain or body if and only if the other is as well. Once again, Conjunctive Causation and Placeless Perceptions together entail that no two similar and simultaneous unit perceptions can differ in their causal relations. For this reason, causal relations to a spatially located brain cannot provide a derivative “place” for perceptions that would circumvent the force of Placeless Perceptions and allow similar and simultaneous unit perceptions to have different causal relations and so, by Causal Bundling, belong to different minds.²¹

The fifth and final unacceptable consequence is the *Failure of Intentionality for Memory*. Because the intentional object of a memory of a perception is determined by its causal relations to a previous perception, there cannot be a subsequent memory of only one of two similar and simultaneous perceptions. For again, Conjunctive Causation and Placeless Perceptions together entail that no two similar and simultaneous unit perceptions can differ in their causal relations. This applies even to memories of the parts of previous

²¹ Strawson (2001) has argued that Hume’s Appendix worries are the result of his recognition of the need for a location or “place of residence” for the operation of the causal mechanisms that account for our ascriptions of identity and simplicity to the mind. He grants that the brain should serve as such a place of residence for Hume but argues that Hume has developed a dialectical situation in which he fails to make appeal to it. The problem I propose is more radical, however, and cannot be resolved by appeal to the brain, for the reason given here. The problem would be resolved, however, if Hume were willing to identify token perceptions with token states of a brain.
successions of perceptions, memories that are supposed to explain the ascription of identity to those successions of perceptions.

To forestall misunderstanding, it should be emphasized that in no case is the problem one of how, from among all the many perceptions that occur in the world, a mind discovers or recognizes which ones belong to it; there is no suggestion that any mind can perceive or remember all of the perceptions that occur, from which it must then somehow make a selection. Rather, the problems concern how, given the three central doctrines, there can be any minds at all, with their own unshared contents, relations to their own unique body, and memory-like ideas that are always of a unique earlier perception. To assume that all of these are unproblematically given is just to assume that many unit perceptions unproblematically share a common “place”—which Hume has denied.

*Satisfying the Criteria.* Why did Hume become dissatisfied with his explanation of personal identity? I suggest that his commitments to Placeless Perceptions and Conjunctive Causation pose insuperable difficulties for Causal Bundling, and that he came to recognize this. While he may not have seen all of the unacceptable consequences just described, it is plausible that he saw or sensed enough of them to realize that his account of the mind as a causal system requires a separate “place” for the perceptions of each mind that his metaphysics did not explain and could not accommodate. The difficulty is a serious one, and more than perplexing enough in its various consequences to constitute a “labyrinth.” The answer thus satisfies the Crisis Criterion.22

22 In fact, the problem proves to be in some ways analogous to the “contradictions” in the explication of the “material world,” to which he compares it at the outset of the Appendix. For the chief problem in the explication of the material world is also a problem about “location”—specifically, the problem of conceiving how bodies are extended once it is granted that they do not possess any qualities resembling the visual and tactile properties by which we conceive them to be extended (THN 1.4.4, “Of the modern philosophy”).
As we have seen, the three central doctrines whose conjunction proves so problematic are all salient in “Of personal identity” and are all salient precisely when Hume, having “loosen’d all our particular perceptions” both from a substantial soul and from perceived “real connexions” with one another, comes to “explain the principles that unite our successive perceptions in our thought or consciousness” in terms of Causal Bundling. The answer thus satisfies the Origin Criterion.

Allowing a substantial soul, were that possible, would of course resolve the difficulty by releasing Hume from both Placeless Perceptions and Causal Bundling; in addition, it would provide at least the prospect of a different approach to explaining the ascription of identity and simplicity to the mind. Indeed, it is a striking fact about the Appendix discussion of personal identity and simplicity that Hume spends the vast majority of it reviewing his reasons for the first Unrenounceable Principle, and only one paragraph reaffirming the second. This focus on the distinctness of perceptions as distinct existences and the absurdity of a thinking substance would not be surprising, however, if he were concerned particularly by the sudden realization that Placeless Perceptions was incompatible with his previous tacit conception of the mind as a kind of place of all its perceptions. Allowing the perception of “real connexions” between distinct perceptions would likewise remove the need to explain the mind’s ascriptions of unity through principles of association in the imagination, but it would also release Hume from Conjunctive Causation, allowing real causal connexions among token perceptions to do the work of bundling perceptions into minds. Thus, the answer also satisfies the Solution Criterion.
Moreover, while the problem is a serious one affecting his entire conception of the mind as a causal system of perceptions, it specifically concerns his account of the mind as it is developed in “Of personal identity,” and it concerns other matters only through its effect on that account. In particular, it does not directly concern the identity or simplicity attributed to things other than the mind. To be sure, Placeless Perceptions and Conjunctive Causation together pose problems for the existence of causal relations among perceptions even without Causal Bundling. But Causal Bundling is motivated in large part by Placeless Perceptions (in particular, by its rejection of a substantial soul), and it is precisely the causal relations among perceptions supposed to be in the same mind that are undermined by the former two doctrines. Thus, the answer satisfies the Scope Criterion.

Finally, while the problem involves the inadequacy of a simple picture of the mind as a place of perceptions, it remains sufficiently hard to describe clearly, and its unacceptable consequences sufficiently difficult to map out, that Hume could understandably have found difficulty in articulating it—while also hoping that an allusion to the sudden “vanishing of his hopes” when “explaining” the uniting principles of mind might be sufficient to indicate its character to his readers. It thus satisfies the Difficulty Criterion. It is a problem that Hume both could have, and should have, seen. If, as seems likely, no other problem yet proposed meets the criteria for a successful answer to the question of his second thoughts, then we should conclude that the explanation of Hume’s second thoughts indeed lies in the direction pointed out to us by Stroud.23

23 I wish to thank the editors as well as Jonathan Cottrell, Peter Kail, Galen Strawson, participants in the 2009 Oxford Seminar in Early Modern Philosophy, and an audience at Boston University for helpful comments.
References


