



**“All Is Revolution in Us”: Personal Identity in Shaftesbury and Hume**

Kenneth P. Winkler

*Hume Studies* Volume XXVI, Number 1 (April, 2000) 3-40.

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](mailto:humestudies-info@humesociety.org)

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

# “All Is Revolution in Us”: Personal Identity in Shaftesbury and Hume

KENNETH P. WINKLER

## 1. Introduction

Even philosophers who believe there is a single “problem of personal identity” conceive of that problem in different ways. They differ not only in their ways of stating the problem, but in the parts of philosophy to which they assign it, and in the resources they feel entitled to call upon in their attempts to deal with it. My topic in this paper is an eighteenth-century uncertainty about the place within philosophy of the problem of personal identity. Is it a problem in metaphysics, or a problem in ethics? Here I try to show that the boundary between ethics and metaphysics was—for a line of philosophers beginning with Locke, continuing with Shaftesbury, and ending (at least in the present paper) with Hume—a shifting and sometimes disputed one. I hope what I have to say will help to clarify a longstanding problem in the interpretation of Hume: his motive for repudiating, in the Appendix to the *Treatise*, the account of judgments of personal identity and simplicity he had provided in Book I.

Whether the topic of personal identity is ethical or metaphysical is a question raised by Locke’s well-known observation that *person* is a “Forensick” or legal term, “appropriating Actions and their Merit; and so belongs only to intelligent Agents capable of a Law, and Happiness and Misery” (*Essay* II xxvii 26).<sup>1</sup> The *Essay*’s treatment of identity, Book II, ch. xxvii, was added to the book in its second edition. Locke had asked his friend William Molyneux whether there were logical or metaphysical topics he had neglected in the first

edition. Molyneux named the principle of individuation, and Locke drafted II xxvii in response. To locate the general topic of identity under the heading of logic or metaphysics is, it would seem, to locate personal identity there as well; this suggests that once we have our account of personal identity in hand, we should be able to determine whether a later person  $y$  is the same as some earlier person  $x$  without considering any science—ethics, for example—not thought to depend on logic or metaphysics. But if *person* is a forensic or legal term, perhaps the judgment that  $y$  is  $x$  can be overthrown by the injustice—the moral inappropriateness—of attributing  $x$ 's acts to  $y$ . If the attribution is inappropriate simply because  $x$  and  $y$  are, according to some purely metaphysical criterion, distinct, the priority of metaphysics over ethics will not, of course, have been compromised. But perhaps Locke has something more dramatic in mind when he speaks of the forensic character of *person*. Perhaps he means to allow that a judgment of personal identity can exhibit a specifically moral shortcoming—a shortcoming to which we have access only by virtue of moral considerations. In that case, judgments of identity will be attendant upon moral considerations. They will not be prior to (and therefore independent of) ethics and practice, as traditional ways of dividing the subfields of philosophy suggest they should be.<sup>3</sup>

The boundary between ethics and metaphysics, or Locke's view of the boundary, has something to do, I think, with how seriously we should take some of the interpretive questions raised by *Essay* II xxvii. Locke tells us that in dealing with questions about identity, we should take notice "what the Word *I* is applied to" (II xxvii 20). Suppose for the moment that materialism is true, and that thinking things are nothing but systems of matter, fitly disposed to serve as vehicles for the superadded attribute of thought. What is the relationship, metaphysically considered, between my self—the person I am—and the living system of matter (the animal or "man") in which my consciousness is (as we would now say) realized? One answer is that the system of matter or animal is a substance and that I am a mixed mode. Another answer is that we are *both* substances, each distinct from the other. Each of these answers raises difficulties, to which Locke's interpreters have responded, often with great ingenuity. But it is possible that Locke never really worried about the questions his interpreters are answering, because he didn't think that an ethically satisfying account of persons called for a foundation in the metaphysics of substance and mode.

In the rest of this paper, I consider Locke only indirectly, in relation to Shaftesbury and Hume. Before turning to them, I should explain what I take to be included in the various subfields of philosophy that figure in the paper. By *metaphysics* I understand the attempt to articulate the nature of things, or the nature of things insofar as we are able to conceive of them, using either the categories of the tradition or categories of one's own invention. By *ethics* I understand the attempt to determine how one should live. I will also be

discussing a third, less traditional subfield within philosophy, "the science of man," by which I understand the attempt to describe or explain, in naturalistic ways, fundamental features of human nature and the human condition. The science of man is of course identified with Hume, but Locke and Shaftesbury are the first to appear in his list of those who have begun "to put the science of man on a new footing" (T xvii).<sup>3</sup>

My presentation will take the following course. I will present Shaftesbury as a critic of Locke inclined to locate personal identity not in the understanding, but in the will. He makes perceptive objections to Locke's account.<sup>4</sup> In the end, though, despite his appeal to the will and even to substance, Shaftesbury does not offer a fully elaborated metaphysics of personal identity. He renounces metaphysics, or declines serious engagement with it, in favor of an appeal to ethics or practice, and to an unanalyzed notion of the self and its identity on which ethics or practice (in his view) depends. I will then turn to Hume, who reinstates metaphysics and makes real improvements over both Shaftesbury and Locke. But in the Appendix to the *Treatise* he faces an explanatory failure—a failure anticipated by Shaftesbury, and one that matters to Hume (or should matter to him) because of its ethical involvements.

## 2. Shaftesbury

Shaftesbury is impatient with "super-speculative Philosophy" or metaphysics, which he dismisses in favor of a philosophy of "a more practical sort" (A 1: 292; C 131).<sup>5</sup> He grants that speculative articulation isn't completely useless, but real philosophy, he contends, should "refine our Spirits, improve our Understandings, or mend our Manners" (A 1: 289; C 129). Defining material and immaterial substances, distinguishing their properties and modes, determining whether an idea is simple or complex—none of these technical pursuits will be, he thinks, of much use to us.<sup>6</sup> "The sum of philosophy," he writes, is "to learn what is just in society and beautiful in Nature and the order of the world." Yet even this will not make one a "real PHILOSOPHER," he thinks, because real philosophy demands "a Heart and Resolution" as well as knowledge (MR 3: 161; C 407). He writes of the importance of observing, "with diligence, what passes *here*," in his mind, "what Connexion and Consistency, what Agreement or Disagreement I find *within*" (A 1: 299–300; C 134). Shaftesbury is seeking the principles by which his judgments are made, his decisions settled, and his tastes formed. If, after identifying these principles, he finds that he is able to endorse them (and Shaftesbury's description of the endorsement process closely resembles Charles Taylor's description of what he calls "strong evaluation"),<sup>7</sup> he will achieve the constancy, unity, or integrity that is a large part of his philosophical aim. This constancy, unity, or integrity is a form of identity or selfhood. It is a selfhood that can only be achieved. It can be achieved, moreover, with more or less success, and it is best achieved

deliberately, with the assistance of what Shaftesbury alternately describes as reason, philosophy, or the moral sense.

Shaftesbury asks himself

whether, according to my present *Ideas*, that which I approve this Hour, I am like to approve as well the next: And in case it be otherwise with me; how or after what manner, I shall relieve my-self; how *ascertain* my *Ideas*, and keep my Opinion, Liking, and Esteem of things, *the same*. If this remains unsolv'd; if I am still the same Mystery to my-self as ever; to what purpose is all this Reasoning and Acuteness? Wherefore do I admire my Philosopher, or study to become such a one, my-self? (A 1: 300; C 134)

Shaftesbury's search for these principles is a model for Hume's more systematic search in the *Treatise of Human Nature*, where Hume gives an inspiring statement of the motive for his philosophizing that recalls this passage from the *Characteristicks*.<sup>8</sup> "I cannot forbear," Hume writes, "having a curiosity to be acquainted with the principles of moral good and evil, the nature and foundation of government, and the cause of those several passions and inclinations, which actuate and govern me." "I am uneasy," he continues, "to think I approve of one object, and disapprove of another; call one thing beautiful, and another deform'd; decide concerning truth and falshood, reason and folly, without knowing upon what principles I proceed" (T 270–271).<sup>9</sup> Shaftesbury had been an earlier victim of the same uneasiness, and he sought to relieve it in roughly the same way.

Shaftesbury directs a number of points against Locke's account of personal identity. They set the stage for his positive view, which is more resolutely independent of metaphysics than Locke's own.<sup>10</sup>

(i) In *The Moralists*, Theocles suggests that "in many things we have been concern'd to purpose, of which we have now no Memory or Consciousness remaining" (M 2: 236; C 254). For all we know, he adds, this may happen "again and again, to perpetuity." Memory, in other words, is not necessary for personal identity. Theocles then continues as follows:

All is *Revolution* in us. We are no more the self-same Matter, or System of Matter, from one day to another. What Succession there may be *hereafter*, we know not; since even *now*, we live by Succession, and only perish and are renew'd. 'Tis in vain we flatter our-selves with the Assurance of our Interest's ending with a certain Shape or Form. What interested us *at first* in it, we know not; any more than how we have *since* held on, and continue *still* concern'd in such an Assemblage of fleeting Particles. (M 2: 236–237; C 254)

This impassioned speech goes beyond criticism of Locke's account of personal identity. Theocles seems to be questioning whether our present body, or any body, is essential to the self. Nothing, he seems to be saying, or at least nothing we can identify, is necessary to us. Philocles, who takes him to be questioning the very substance or being of the self, later on develops an ethical response. "As accidental as my Life may be," he says, "or as that random Humour is, which governs it; I know nothing, after all, so *real* or *substantial* as MY-SELF. Therefore if there be that Thing you call a *Substance*, I take for granted I am one. But for anything further relating to this Question, you know my *Sceptick* Principles: I determine neither way" (M 2: 353–354; C 302).

(ii) A second criticism of Locke appears in the *Miscellaneous Reflections*, where Shaftesbury asks "what constitutes the WE or I," and "whether the I of this instant, be the same with that of any instant preceding, or to come?" (MR 3: 193; C 420). We have, he continues,

nothing but *Memory* to warrant us: And *Memory* may be false. We may believe we have thought and reflected thus or thus: But we may be mistaken. We may be conscious of that, as *Truth*; which perhaps was no more than *Dream*: and we may be conscious of that as a *past* *Dream*, which perhaps was never before so much as dreamt of.

This is what *Metaphysicians* mean, when they say, "that *Identity* can be prov'd only by *Consciousness*; but that *Consciousness* withal, may be as well false as real in respect of what is past." So that the same successional *We* or *I* must remain still, on this account, undecided. (MR 3: 193–194; C 420–421)

The upshot of this objection is that memory is not *sufficient* for personal identity. The author of the *Miscellaneous Reflections* represents himself as a commentator on the treatises that come before it—a commentator distinct from their author. In response to his own objection, the commentator, like Philocles in the *Moralists*, proposes to take his own being "upon Trust" (MR 3: 194; C 421). "Let others philosophize as they are able: I shall admire their strength, when, upon this Topick, they have refuted what able *Metaphysicians* object, and PYRRHONISTS plead in their own behalf." "Mean while," he continues,

there is no Impediment, Hindrance, or Suspension of *Action*, on account of these wonderfully refin'd *Speculations*. Argument and Debate go on, still. Conduct is settled. Rules and Measures are given out, and receiv'd. Nor do we scruple to act as resolutely upon the mere Supposition that *we are*, as if we had effectually prov'd it a thousand times, to the full satisfaction of our *metaphysical* or *Pyrrhonian* antagonist. This to me appears sufficient Ground for a *Moralist*. Nor

do I ask more, when I undertake to prove the reality of VIRTUE and MORALS.

In its appeal to conduct or practice, this response resembles Berkeley's response in a similar context in *Alciphron*.<sup>11</sup> It also resembles Locke, who favored his account of personal identity at least partly because it allowed him to pry pressing questions of conduct away from traditional metaphysical issues—for example, is the thinking self material or immaterial?—on which they had been thought to depend. But the most striking resemblance is to Hume, whose response to excessive Pyrrhonism Shaftesbury here anticipates.<sup>12</sup>

(iii) Shaftesbury's ethical orientation—his emphasis on concernment or concern (see A 1: 294, 302; C 131–132, 135)—is, then, reminiscent of Locke.<sup>13</sup> But in *Advice to an Author* Shaftesbury expressly rejects Locke's account, on the basis of a distinctly "ethical" thought experiment. He invites us to imagine someone who has undergone a complete "Revolution" in his passions, affections, and opinions (A 1: 285; C 127). He claims that we would not "attempt any renewal of Acquaintance or Correspondence" with such a person, "tho perhaps he might preserve in his Memory the faint Marks or Tokens of former Transactions which had pass'd between us" (A 1: 284; C 127). Memory or sameness of consciousness is, once again, insufficient. Lockean identity is too thin to be ethically decisive.

(iv) In the *Philosophical Regimen* Shaftesbury raises a closely related objection:<sup>14</sup>

The metaphysicians and notable reasoners about the nice matters of identity, affirm that if memory be taken away, the self is lost. And what matter for memory? What have I to do with that part? If, *whilst I am*, I am but as I should be, what do I care more? and thus let me lose *self* every hour, and be twenty successive selves, or new selves, 'tis all one to me: so I lose not my opinion. If I carry that with me 'tis I; all is well.

This true opinion is of the self "as not body" even "whilst in a body."<sup>15</sup> A mind, he writes,

is something that acts upon a body; and not on a body only, but on the senses of a body, the appearances, fancies, and imaginations, by correcting, working, modelling these, and building out of these. Such is a mind. Such a thing I know there is in the world somewhere. Such a mind I am sure of. Let Pyrrho by the help of such a mind contradict this if he please. He and I have each of us our individual understandings.<sup>16</sup>

The mind of which Shaftesbury speaks is the seat of the active principles—principles of judgment, resolution, and taste—that he hopes to identify and endorse. The identity of the mind consists not in memory or consciousness, but in these. As the seat of principles meant to govern or rule, the mind, in this essential aspect, can be described as the will. Political analogies or metaphors of mind are, of course, at least as old as Plato, and when Shaftesbury writes that "all is revolution in us," his comment can be read in two ways. He may simply be reporting the incessant change within us, as Hume does at *Treatise* I i 4, where he speaks (T 11) of the "constant revolution of our ideas." But Shaftesbury may also be expressing a fear of personal disintegration, a loss or absence of the kind of governance that complete selfhood requires. A complete self needs a constitution. Otherwise it threatens to become

*a Wilderness; where all is laid waste, every thing fair and goodly remov'd, and nothing extant but what is savage and deform'd. Now if Banishment from one's Country, Removal to a foreign Place, or any thing that looks like Solitude or Desertion, be so heavy to endure; what must it be to feel this inward Banishment, this real Estrangement from human Commerce; and to be after this manner in a Desart, and in the horridest of Solitudes, even when in the midst of Society? What must it be to live in this Disagreement with every thing, this Irreconcilableness and Opposition to the Order and Government of the Universe? (I 2: 171; C 229)*

In this passage, Shaftesbury speaks of estrangement from the deepest principles of his mind. It invites comparison to the concluding section of Book I of the *Treatise*, where Hume struggles with his own alienation and solitude, and speaks, as Shaftesbury does, of his isolation from "human commerce" (T 264). One crucial difference is that Shaftesbury's principles are, like the *logos* of the Stoics, built into reality itself; the principles from which Hume suffers estrangement have no such cosmic backing.

Despite these criticisms of Locke, Shaftesbury actually embraces—though he also revises—what might be called the "skeptical structure" of Locke's account. That structure has five main elements:

- (i) a distinction between appearance and underlying reality (the appearance, in Locke's case, being consciousness or memory, and the underlying reality being the substances—"whether Spiritual, or Material, Simple, or Compounded" (*Essay* II xxvii 17)—in which consciousness is realized);
- (ii) the claim that we are condemned to ignorance of the underlying reality (thus Locke speaks at *Essay* II xxvii 21 of "this ignorance we are in of the Nature of that thinking thing, that is in us");

(iii) the claim that the appearance is transparent, or nearly so (thus Locke, in describing the identity or diversity of substance as something “*which [we] cannot be sure of*” (*Essay* II xxvii 25), implies that the scope of consciousness is something of which we can be sure, though there remains the problem, addressed for example at II xxvii 22, of adjusting attributions of responsibility made from the third-person point of view to those made from the viewpoint of the self);

(iv) the assurance that the appearance is ethical or practically sufficient (or more precisely, that the appearance, once appreciated, provides ethics or practice with all that philosophy could ever provide);<sup>17</sup>

(v) the assurance that the appearance is also sufficient for religious faith, or for faith insofar as it can be supported by philosophy.

Locke’s recognition of the ethical or practical sufficiency of the appearance is most apparent in the ethical emphasis of *Essay* II xxvii 18–20 and 26, and his assurance of religious sufficiency is clearest at *Essay* II xxvii 15, where we are told that with Locke’s account in hand “we may be able without any difficulty to conceive, the same Person at the Resurrection.” In speaking of the “skeptical structure” of *Essay* II xxvii, I do not mean to suggest that Locke himself regarded it as skeptical. But as we will see, Shaftesbury’s revision of the structure is avowedly skeptical, and Locke’s discussion resembles Shaftesbury’s—and skeptical thinking more generally—in its renunciation of knowledge for the sake of life or practice.

Shaftesbury follows Locke in distinguishing between appearance and underlying reality. Like Locke, he renounces knowledge of the reality, and like Locke again, he reassuringly insists that what we find in the appearance is enough to meet our needs. Locke’s great mistake, in Shaftesbury’s view, was to give the appearance of consciousness an explanatory role—a role it could not fill, as Shaftesbury’s criticisms show. The consciousness of which Locke speaks is not in fact transparent, or if it is, it cannot do all the work he asks of it. Whatever his intentions, Locke’s consciousness becomes another underlying reality, even if it does not lie as deep as the substances—spiritual or material, simple or compound—that God or angels might be able to see beneath it. Locke becomes a metaphysician despite himself. Shaftesbury, by contrast, remains bravely at the level of appearance.<sup>18</sup> As I will explain below, he does infer an inner principle that makes us one, building on doubt-free “foundations . . . taken from our very *Perceptions, Fancys, Appearances, Affections* and *Opinions* themselves, without regard to any thing of an *exteriour WORLD*” (MR 3: 212; C 428). But he claims to know nothing about that principle as it is in itself—he has, to borrow some words that Locke made famous in another context, “no *Idea* of what it is, but only a confused obscure one of what it does” (*Essay* II xiii 19)—and he attempts no account of how either the principle or

the appearance it underlies supports our life needs. Shaftesbury feels, I think, that to address these questions in an explanatory spirit is to engage in metaphysics, and to expose oneself to the kind of objections he makes against Locke. Locke uses the appearance he calls "consciousness" to account for our retributive practices, and to secure robust theological commitments, among them the possibility that we will one day be resurrected. Shaftesbury's theology carries no such commitments; his immanent God resembles a Platonic world-soul, or the *logos* of the Stoics, far more than it resembles the Biblical creator who imposes laws and enforces them. As Darwall observes, Shaftesbury "holds no brief for the notions of warranted blame or punishment."<sup>19</sup> This makes it easier for him to remain at the surface, or very close to it, with an unadorned and arguably unearned assurance that we are both substantial and self-identical. Thus Philocles "take[s] for granted" that he is a single substance (M 2: 353; C 302), and sets further questions to one side by pleading what Theocles calls, in a passage to which we will return, the "Privilege of *Scepticism*" (M 2: 354; C 302). The author of the *Miscellaneous Reflections* follows suit. How we are "continu'd *one and the same*" over time, he admits, "is not a Matter . . . easily or hastily decided" (MR 3: 192; C 420). "For my own part," he says, "I take my Being upon Trust" (MR 3: 194; C 421). The only reward of metaphysics is the knowledge that "there is no *Knowledg* or *Wisdom*" to be gained in this part of philosophy (MR 3: 210–211; C 427); the more quickly we pass over it, the better.

I turn now to Shaftesbury's wider views. A main theme in Shaftesbury's positive account of identity, one especially prominent in *Advice to an Author*, is the power we have of dividing ourselves into "two Partys" (A 1: 169; C 77). To know ourselves we must divide ourselves in two (A 1: 170; C 77), so that one part can contemplate and evaluate the other. This process of contemplation and evaluation is what he calls the "Method of SOLILOQUY" (A 1: 160; C 73 and elsewhere). Self-knowledge or self-comprehension is constitutive of our identity: the known province of philosophy is "to teach us *our-selves*, keep us the *self-same* Persons, and so regulate our governing Fancys, Passions, and Humours, as to make us comprehensible to our-selves, and knowable by other Features than those of a bare Countenance" (A 1: 283; C 127). Shaftesbury seems to think that we won't be self-comprehensible if we are at the mercy of our passions. Lacking what he calls a stable "aim" or "end" (A 1: 308, 307; C 138, 137), we won't be able to construct a compelling narrative of our lives. We will therefore lack a "real SELF" (A 1: 307; C 137). Because I can judge all of my fancies, I stand apart from all of them. "FANCY and I are not *all one*. The Disagreement makes me *my own*" (A 1: 325; C 145). The real or true self is the controller or manager—the rational self who rules. An aim is described as a "*Resolution*," and it is attributed to the will—a rationally informed will that makes me "*one and the same* Person to day as yesterday, and to morrow as to day" (A 1: 187; C 84).

It is tempting to say that all this has nothing to do with the classical philosophical problem of personal identity, a problem about the conditions for a thing's persistence over time. And perhaps there is no interesting connection between that problem and the challenge of personal integrity or unity as Shaftesbury understands it, even when it presents us with a task to be discharged over time.<sup>20</sup> Shaftesbury seems to be concerned with what Derek Parfit calls "what matters," and *his* what matters isn't confined to "what matters in identity."<sup>21</sup> It is stretched far beyond that. The fact is that what I care about in *x* may have nothing to do with *x*'s identity in the numerical sense. What I want *x* to achieve, or what I want to preserve in *x*, may not be what makes *x* *x*.

Shaftesbury himself actually raises a related objection. Theocles argues that Philocles could survive a change in even his deepest principles. If you, Philocles, were to become convinced of my argument for "the Divine Hypothesis," says Theocles,

and from this new Turn of Thought admit a total Change in all your Principles and Opinions; yet wou'd you still be the self-same PHILO-CLES: tho better yet, if you will take my Judgment, than the present-one, as much as I love and value him. You see therefore, there is a strange Simplicity in this YOU and ME, that in reality they shou'd be still *one and the same*, when neither *one* Atom of Body, *one* Passion, nor *one* Thought remains the same. (M 2: 351; C 301)

Shaftesbury's own strategy for the conduct of self-assessment—a dividing of the self into two, one of which can, it seems, make anything in the other an alien object of evaluation—can press us toward the conclusion that the self is indeed "strangely simple." Shaftesbury tells us that we can divide ourselves off from virtually everything in us in order to decide whether to make it our own, or make ourselves such as to include it. The self is in danger of becoming "punctual."<sup>22</sup> At T 261, Hume suggests that we can survive *any* change, including the dramatic change in "character and disposition" that Theocles contemplates.<sup>23</sup> Does Shaftesbury agree? And if he does, must he embrace the hypothesis of "strange Simplicity" that Theocles advances?

I do not think we can safely conclude that Shaftesbury himself agrees with Hume, or that he accepts our "strange"—by which Theocles probably means our *great* or *extreme*—simplicity.<sup>24</sup> One obstacle to associating Shaftesbury with Theocles is the dialogue form of *The Moralists*.<sup>25</sup> Philocles describes his friend as an "Enthusiast"—"let me hear a-new," he says, "that Divine Song with which I was lately charm'd" (M 2: 359; C 304)—and if Shaftesbury agrees, he may stand, with Philocles, at some distance from Theocles, even if he finds himself moved by what Theocles has to say.<sup>26</sup> A second obstacle is that the premise Theocles uses to defend his hypothesis falls far short of its conclusion. If in the end Philocles does convert to the immanent theism of Theocles, it may be

because of enduring commitments deeper than either theism or atheism—commitments that secure his identity even if their depth frustrates his attempts to articulate them. These commitments may be what Theocles most values in his friend, even if he cannot spell them out. Third, some of Shaftesbury's own arguments against Locke run counter to the hypothesis of strange simplicity: why would we want to renew our acquaintance with someone who retained nothing of the character we cherished, simply because the earlier and later persons had some atom of selfhood in common? Fourth and finally, the highly integrated identity Shaftesbury is seeking is not strangely simple but richly articulated: it would be altogether lost if our founding commitments were to shift.

The fourth consideration at least is not decisive. Shaftesbury believes that the identity he seeks is *set all of us as a task*, a task that presumably presents itself to those (of "us") whose identities will one day be (it is hoped) more fully achieved. Perhaps the selves to which this task is set *are* strangely simple, even if their more integrated counterparts—our perfected selves—are not. There seem, then, to be two selves at work in Shaftesbury's texts—the highly integrated, reflectively endorsable self he urges each of us to become, and the less collected but nonetheless active self he takes himself to be addressing—and the to-be-perfected self may be strangely simple even if the perfected self is not.<sup>27</sup>

It does not follow, of course, that the self to-be-perfected—the self of Shaftesbury's audience—*must* be strangely simple. The perfected self could be constituted by principles that descend, by a process that renders them consistent and self-conscious, from the principles that constitute the self to-be-perfected. And although he offers only hints as to how to this process might be conducted, it seems to me that this is what Shaftesbury really has in mind. He may not agree with Theocles that I am strangely simple. But even if he does, he may not interpret simplicity as absolute unanalyzability (or unanalyzability in principle). Theocles offers the hypothesis of strange simplicity in the context of a wider discussion, in which he considers the unity of plants, animals, and (as we will see in a moment) the cosmos as a whole. He holds that every animal or vegetable has what he calls a "peculiar *Nature*" (M 2: 348–349; C 300). This nature is not unanalyzable in principle: it is that "by virtue" of which a tree, for example, is "a real *Tree*," able to live, flourish, and remain "still *One and the same*; even when by Vegetation and Change of Substance, not one Particle in it remains *the same*" (M 2: 349; C 300). Despite the complex role he assigns to them, Theocles is willing to describe these natures as "simple" (M 2: 352; C 301); given the fact that his remarks on the self—"the thing of *Personality* between you and me" (M 2: 350, C 300)—are closely modeled on what he says about these natures (see M 2: 350–352; C 300–301), it is reasonable to conclude that our own "simplicity" is compatible with a kind of complexity: a complexity that distinguishes our kind from other kinds, and

individuals within our kind from other individuals, even if we are unable to say what it consists in.<sup>28</sup>

Shaftesbury's final view, then, may be that certain governing principles, ascribable to the will, are essential to us, even in our to-be-perfected state, but neither Theocles nor Philocles even attempts to explain how.<sup>29</sup> Philocles, in fact, pleads what Theocles calls "the . . . Privilege of *Scepticism*" (M 2: 354 and C 302; the privilege is also mentioned, in another context, at M 2: 208 and C 242): the privilege to go on living and believing as before, when reason is unable to determine the judgment in either of two ways. The invocation of this privilege has to strike any reader of Hume, who pleads the privilege of a skeptic in the Appendix to the *Treatise*.

Shaftesbury, as I have already noted, does not confine his attention to the identity of persons, and his remarks on the identity of natural objects—including the cosmos as a whole—provoked a direct response from Hume. In the *Treatise*, when Hume begins his account of our belief in identity, he attributes that belief in identity to our propensity to confuse identity with close relation, particularly resemblance. We

boldly assert that these different related objects are in effect the same, however interrupted and variable. In order to justify to ourselves this absurdity, we often feign some new and unintelligible principle, that connects the objects together, and prevents their interruption or variation. Thus we feign the continu'd existence of the perceptions of our senses, to remove the interruption; and run into the notion of a *soul*, and *self*, and *substance*, to disguise the variation. But we may farther observe, that where we do not give rise to such a fiction, our propensity to confound identity with relation is so great, that we are apt to imagine something unknown and mysterious, connecting the parts, beside their relation; and this I take to be the case with regard to the identity we ascribe to plants and vegetables. (T 254)

For an illustration of "how a great genius may be influenc'd by these seemingly trivial principles of the imagination, as well as the mere vulgar," a footnote directs the reader to "Lord *Shaftesbury's* reasonings concerning the uniting principle of the universe, and the identity of plants and animals. See his *Moralists*: or, *Philosophical rhapsody*" (T 254).

For the kind of passage Hume has in mind we can turn to one in which Theocles invites Philocles to infer from the fact that the world is one—that it "hangs together, as *of a Piece*"—that "there shou'd be something belonging to it which makes it *One*" (M 2: 347; C 299). On this basis, Theocles puts forward a general principle: "wherever there was such a *Sympathizing of Parts*, as we saw here, in our *real TREE*; Wherever there was such a plain *Concurrence in one common End*, and to the Support, Nourishment, and Propagation of so fair a

*Form*; we cou'd not be mistaken in saying there was a peculiar *Nature* belonging to this *Form*, and common to it with others of the same kind" (M 2: 348–349; C 300). Philocles doubtfully replies that if this argument is accepted, we will be left with a whole store "of immaterial and immortal Substances" (M 2: 349; C 300). Theocles, in reply, refuses to insist on immortality, but the burgeoning population of immaterial natures or substances leaves him undisturbed.

I will return to this passage (and others like it) later on. Insofar as he feigns imaginary principles of union, Shaftesbury may be open to Humean criticism. But Shaftesbury raises a question that returns, in the Appendix to the *Treatise*, to haunt Hume. Shaftesbury has Theocles say:

When you have argu'd ever so long upon these Metaphysical Points of *Mode* and *Substance*, and have philosophically concluded from the Difficultys of each Hypothesis, "That there cannot be in Nature such a *Universal-One* as This;" you must conclude, from the same Reasons, "That there cannot be any such *particular-One* as Your-self." But that there is actually such a one as *this latter*, your own Mind, 'tis hop'd, may satisfy you. (M 2: 354; C 302)

Elsewhere:

Is not this *Nature* still a SELF? Or, tell me, I beseech you, How are You *one*? By what *Token*? or by virtue of *What*? "By a Principle which joins certain Parts, and which thinks and acts consonantly for the Use and Purpose of those Parts." Say, therefore, What is your whole System a Part of? Or is it, indeed, no *Part*, but a *Whole*, by *it-self*, absolute, independent, and unrelated to any thing besides. If it be indeed a *Part*, and really *related*; to what, I beseech you, but to *the Whole of Nature*? Is there then such a *uniting Principle* in NATURE? If so, how are you then a *Self*, and *Nature* not so? (M 2: 356–357; C 303)

In what sense, Shaftesbury asks, am I one and nature not? I think this question spells trouble for Hume.

### 3. Hume

I propose to concentrate less on Hume's account of personal identity in *Treatise* I iv 6 than on his four pages of second thoughts in the Appendix. I hope to explain in a new—or at least newly convincing—way why Hume became dissatisfied with his account of our judgments of personal identity and simplicity.

One of Hume's ambitions in the *Treatise* was to outdo Locke, the first person on Hume's list of "late philosophers in *England*, who have begun to put the

science of man on a new footing" (T xvii). Locke offered the public an essay on human understanding; Hume's offering is both more confident—a treatise rather than an essay—and broader in scope. The understanding, the announced concern of all of Locke's book, is dealt with in just the first volume of Hume's, which is subtitled "Of the Understanding."<sup>30</sup> Locke's book is not a traditional treatise on metaphysics; it takes up metaphysical topics insofar as our ideas (whose origin and content Locke aimed to describe) occasioned them. The *Essay* does, however, satisfy the definition of metaphysics I gave when I began, which includes the attempt to articulate the nature of things insofar as we are able to conceive of them. In *Essay* II xxvii, "Of Identity and Diversity," Locke was concerned above all with a forensic notion, and he hoped to pry important ethical questions away from their alleged metaphysical foundations. Hume is also interested in articulating our ideas and tracing their origin in experience. But unlike Locke, in *Treatise* I iv 6, "Of personal identity," he offers metaphysics to the reader and makes no excuse for it.

Hume is, like Locke, a metaphysician of things insofar as we are able to conceive of them. The basic unit of his metaphysics of the person is not the substance but the *perception*. There are two kinds of perceptions, differing only in "the degrees of force and liveliness, with which they strike upon the mind": the more lively are *impressions*, the less lively *ideas* (T 1). The "true idea of the human mind," he announces in I iv 6, "is to consider it as a system" of such perceptions, "link'd together by the relation of cause and effect" (T 261). One perception produces another and is later "expell'd" by a third (T 261). "In this respect," Hume writes,

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. Whatever changes he endures, his several parts are still connected by the relation of causation. (T 261)

Elsewhere in *Treatise* I iv 6 Hume compares the mind to a theater, "where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations" (T 253). "The comparison of the theatre," he cautions, "must not mislead us. They are the successive perceptions only, that constitute the mind." We are nothing "but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement" (T 252).

In these various passages, Hume is telling us what the self (insofar as we are able to conceive or understand it) *is*. Unlike Locke, who believes that substance of some sort (whether spiritual or material, simple or compound) underlies the self, even if it cannot be known, Hume has no notion at all of substance, or no notion that distinguishes it from a perception (T 233). A reader of Locke might be tempted to think that the underlying substantial basis of consciousness is the true or real self. Hume permits no such temptation. He altogether dismisses the question—*is self the same with substance?*—that had occupied Locke in *Essay* II xxvii 12–14. “For my part,” Hume writes, “I have a notion of neither [self nor substance], when conceiv’d distinct from particular perceptions” (T, Appendix, 635).

In claiming to identify “the true idea of the human mind,” Hume is in one way more ambitious than Locke, whose belief in a substantial basis for consciousness prevents him from speaking with such confidence. But in another way Hume is less ambitious, because Hume in *Treatise* I iv 6 is considering only the mind or understanding, whereas Locke, in *Essay* II xxvii, despite his book’s more modest title, was bringing the whole person into view. At T 253, Hume insists that we “distinguish betwixt personal identity, as it regards our thought or imagination, and as it regards our passions or the concern we take in ourselves.” Only the first, he says, is at issue in I iv 6. Book I of the *Treatise* is “Of the Understanding,” and Hume presses the distinction between thought and imagination on the one hand and passion and concern on the other because, in Book I, he intends to leave the body out of account. For Locke in the *Essay*, the body was always a part—a literal part—of the self or person. “Our very Bodies,” Locke writes in *Essay* II xxvii 11, “are a part of our selves: *i.e.* of our thinking conscious self.” Thus “the Limbs of his Body is to every one a part of himself. He sympathizes and is concerned for them.”<sup>31</sup> In the second book of the *Treatise*, “Of the Passions,” where Hume himself takes up our self-concern and our passions, the body is restored to the self, as he indicates (having raised the question of the body’s status at T 298) at T 303.<sup>32</sup>

Shaftesbury was unhappy with Locke’s account of personal identity. The lesson he might have drawn from Locke’s failure is that it is a mistake to seek the mind’s identity in any of its faculties, but Shaftesbury concluded that Locke had fastened on the wrong one: understanding (or consciousness) rather than will. Hume is by no means insensitive to the difference between personal identity with regard to the understanding (“as it regards our thought or imagination,” T 253) and personal identity with regard to the will (“as it regards our passions or the concern we take in ourselves,” T 253). But he does not think that our identity consists in the persistence of *any* faculty. Nor does he find any use for the appeal to substance. Hume’s metaphysics of personal identity is both “anti-facultative” and “anti-substantialist.” (It is “reductionist” in Derek Parfit’s sense, and “relational” in Raymond Martin’s sense.)<sup>33</sup> What makes us the same through time, insofar as we are the same through time, are

relations among successive perceptions, particularly *causal* relations. This enables Hume to make many improvements on his predecessors. But it seems to me that it leads him into a difficulty—and, in the Appendix, into recognizing a difficulty—that Shaftesbury foreshadows.

Hume in the Appendix remained happy with the metaphysics of I iv 6; he announces, after reviewing that metaphysics, that “so far I seem to be attended with sufficient evidence” (T 635). The failure he admits is an *explanatory* failure—a failure in his science of man, rather than in his metaphysics. This is a point many others have made, but some of the evidence in its favor is worth reviewing.<sup>34</sup> After the introductory sentences I have already discussed, Hume reviews the arguments that induced him “to deny the strict and proper identity and simplicity of a self or thinking being” (T 633). These arguments occupy the next nine paragraphs, the last of which announces that just as philosophers have grown accustomed to thinking that we have no idea of external substance distinct from the ideas of particular qualities, they should be receptive to “a like principle with regard to the mind, *that we have no notion of it, distinct from the particular perceptions*” (T 635). “So far,” he then writes, “I seem to be attended with sufficient evidence,” itself a strong indication that the metaphysics of I iv 6 is not at fault.<sup>35</sup> Hume then writes,

but having thus loosen’d all our particular perceptions [that loosening being the upshot of the true metaphysics], when I proceed to explain the principle of connexion, which binds them together, and makes us attribute to them a real simplicity and identity; I am sensible, that my account is very defective, and that nothing but the seeming evidence of the precedent reasonings cou’d have induc’d me to receive it. (T 635)<sup>36</sup>

There is a footnote attached to the word “when”—“when I proceed to explain the principle of connexion, which binds them together”—and it directs us to the first paragraph that begins on T 260, where Hume, having concluded that “our notions of personal identity, proceed entirely from the smooth and uninterrupted progress of the thought along a train of connected ideas,” writes that

the only question . . . which remains, is, by what relations this uninterrupted progress of our thought is produc’d, when we consider the successive existence of a mind or thinking person. And here ’tis evident we must confine ourselves to resemblance and causation, and must drop contiguity, which has little or no influence in the presence case.<sup>37</sup>

All this indicates that what troubled Hume in the Appendix is not his metaphysics of the person, but his scientific explanation of attributions of personal

identity. We can, in fact, pinpoint his dissatisfaction even more precisely. Hume not only affirms his earlier metaphysics, but endorses some of the inferences earlier drawn from it. He continues to believe that it leaves us with the question whether, in judging a person's identity, "we observe some real bond among his perceptions, or only feel one among the ideas we form of them" (T 259). And he continues to think that the question can easily be decided, if we recall "what has been already prov'd at large, that the understanding never observes any real connexion among objects," from which it follows that identity "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). The failure to which he confesses, then, is a failure to identify the relations among perceptions that cause this union in the imagination.

In the Appendix, Hume opens his remarks on personal identity by reporting that he "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" (T 633).<sup>38</sup> But "upon a more strict review," he continues, "I find myself involv'd in such a labyrinth, that, I must confess, I neither know how to correct my former opinions, nor how to render them consistent." In one way, Hume is preparing us for something discontinuous with what we have already seen—a reversal of the optimism of *Treatise* I iv 6, "Of personal identity." But in another way, it seems that we are being prepared for something continuous with what we have already seen—not with the optimism of I iv 6, but with the difficulties of I iv 7, the conclusion of Book I. It was in I iv 7 that Hume surveyed the contradictions and absurdities that vex human reason when it confronts the material world. The opening words of Hume's Appendix make us realize, perhaps for the first time, that contradictions and absurdities in the intellectual world (supposing they exist at all) were, in I iv 7, passed over in silence, despite the close attention that world received in I iv 5 and 6. There is one allusion to sections 5 and 6 in section 7, but it suggests that one of his main contentions in the earlier sections—that a person is merely a succession of perceptions—is entirely secure. Explaining that were it not for habit, we could not attribute an independent existence to the objects of the senses, Hume writes that, habit aside, we would be forced to comprehend or include them "entirely in that succession of perceptions, which constitutes our self or person" (T 265). Despite the doubts and difficulties of I iv 7, then, something in I iv seems to be safe from criticism.

Perhaps it is no surprise that Hume should say his theory of the person is secure. He had claimed, after all, to identify the "true idea of the human mind" (T 261). But it is nonetheless striking that difficulties in understanding the material world should occupy all of Hume's attention in I iv 7. That gives us a reason for being puzzled by the Appendix. If we take him at his word, Hume was at one time confident of his theory of the intellectual world, at least up to

a point: confident that it was free from contradiction and absurdity. Why was he confident, and why, a short time later, did his confidence vanish? And if, as my comments so far may suggest, the aim of the Appendix is to give the difficulties of the intellectual world their due, why is the Appendix so different in tone or feeling from the conclusion of Book I? Why doesn't Hume respond to the difficulties acknowledged in the Appendix in just the way he responds to the supposedly similar difficulties generated by I iv 2, 3, and 4? Why doesn't his skepticism play the same role in the Appendix that it does in I iv 7?

I hope to answer at least some of these questions, and I propose to do so by paying some attention to *Treatise* I iv 2, Hume's account of our belief in the continued and distinct existence of body. Even if my answers to the questions posed are not in the end persuasive, I recommend viewing I iv 6 and the Appendix in this context. There are numerous points of contact. To mention just a few of the more obvious ones, a crucial element in Hume's account of our belief in body—our tendency to find continued existence in interrupted perceptions—is borrowed from what Hume represents as the fuller account of the tendency in I iv 6 (T 206 in I iv 2). And in I iv 2, Hume looks briefly ahead to the theory of the person developed in I iv 6. "We may observe," Hume writes there, "that what we call a *mind*, is nothing but a heap or collection of different perceptions, united together by certain relations, and suppos'd, tho' falsely, to be endow'd with a perfect simplicity and identity" (T 207).

One feature of the Appendix that has troubled very many readers is Hume's apparent claim that the two principles identified at T 636—the distinct existence of distinct perceptions, and the absence of perceived connections among distinct existences—are inconsistent. They are in fact not inconsistent,<sup>39</sup> and if they were, a lot more than I iv 6—I iv 5, for example, in which both principles are prominent—would be compromised.<sup>40</sup> But if Hume, by the time of the Appendix at least, is trying to assimilate the difficulties in his account of personal identity to those discussed in the conclusion of Book I, it isn't surprising that he should speak of "contradictions" (for example at both T 633 and T 636), having made them his concern in I iv 7 (for example at T 266). In other words, Hume may have been inclined, for reasons of structure or closure, to cry contradiction when no contradiction was in view, or at least not in close view. I return briefly to this point in closing.

My explanation of Hume's difficulty, very roughly put, is as follows. Hume's account of personal identity depends crucially on the relation of causation. Resemblance plays a minor or subsidiary role. The problem is that causation is simply unable to generate the kind of "easy transitions" that, in the case of resembling perceptions, enable the mind to ignore (or ignore the real import of) perceived differences. This is especially true of the mind of the philosopher, which has been disciplined to accept the distinctness (or perceived distinctness) of causally related objects—their distinctness from one another, as well as their distinctness from any alleged support or substratum—

a discipline that begins in part iii of Book I but reaches a climax in sections 5 and 6 of part iv. It is hard to see how anyone, especially a philosopher who has felt the force of Hume's earlier arguments, can accept the simplicity and identity of the self.<sup>41</sup>

The first text I want to offer in support of my proposal is T 254, where Hume, in I iv 6, begins his full account of the mind's tendency to identify interrupted and variable objects (the account to which I iv 2 refers us). This passage (which I quoted in my remarks on Shaftesbury) comes before Hume's account of personal identity; it is part of his account of the identity of bodies, among them vegetables, animals, and human beings. He admits that the mind can, for a time, hold off the tendency to take distinct but closely related objects to be the same. "Tho' we incessantly correct ourselves by reflexion, and return to a more accurate method of thinking," he writes, "yet we cannot long sustain our philosophy, or take off this bias from the imagination." Eventually we yield to it, "and boldly assert that these different related objects are in effect the same, however interrupted and variable." But the mind cannot rest with this bold assertion. It feels a need to "justify . . . this absurdity"—its earlier philosophizing has apparently left it with a bad conscience—and therefore feigns

some new and unintelligible principle, that connects the objects together, and prevents their interruption or variation. Thus we feign the continu'd existence of the perceptions of our senses, to remove the interruption; and run into the notion of a *soul*, and *self*, and *substance*, to disguise the variation. But we may farther observe, that where we do not give rise to such a fiction, our propension to confound identity with relation is so great, that we are apt to imagine something unknown and mysterious [it is here that Hume, in a footnote attached to the word "imagine," refers to Shaftesbury], connecting the parts, beside their relation; and this I take to be the case with regard to the identity we ascribe to plants and vegetables. (T 254–255).<sup>42</sup>

The passage distinguishes among the following four tendencies:

- (i) a tendency to suppose that distinct but related objects or perceptions are the same;
- (ii) a tendency to "feign" the continued existence of such objects or perceptions;
- (iii) a tendency to "run into" the notions of soul, self, and substance;<sup>43</sup>
- (iv) a tendency to "imagine something unknown and mysterious" connecting related perceptions or objects—something *besides* their relation.

The four tendencies have a number of notable features. First, of the two relations at work in tendency (i)—resemblance and causation—resemblance is by far the more important. As Hume explains at *Treatise* I iv 2, resemblance is of all relations the one “most efficacious” in causing us to mistake one idea for another, “and that because it not only causes an association of ideas, but also of dispositions, and makes us conceive the one idea by an act or operation of the mind, similar to that by which we conceive the other” (T 203–204).<sup>44</sup>

Second, tendency (ii) is not on a par with tendency (i); it is triggered only after (i) has done its work. Tendency (ii) is an expression of our contradiction-sensitivity, even when we are under the control of what Hume calls imagination. When I identify an earlier object with a later one (having turned my attention elsewhere for a time), I must suppose that what I now take to be one object continued to exist when I did not perceive it. If I fail to make this supposition, I am, in Hume’s view, taking the two perceptions to be distinct, which contradicts the work of (i).

Third, Hume sees a difference of some kind between tendency (i) and tendencies (ii), (iii), and (iv). He writes at T 255 that “the controversy concerning identity is not merely a dispute of words. For when we attribute identity, in an improper sense, to variable or interrupted objects, our mistake is not confin’d to the expression, but is commonly attended with a fiction, either of something invariable or uninterrupted, or of something mysterious and inexplicable, or at least with a propensity to such fictions.” And when he announces later in the section that all disputes concerning the identity of connected objects are merely verbal, he adds: “except so far as the relation of parts gives rise to some fiction or imaginary principle of union, as we have already observ’d” (T 262).

These passages suggest that Hume is more suspicious of (ii), (iii), and (iv) than he is of (i). But what exactly is the difference between them? The passages just quoted suggest that (i) commits us only to *verbal* errors, while the errors involved in (ii), (iii), and (iv) are *substantive* or philosophical. But this cannot be right, because (i) is elsewhere implicated in substantive error, for example at I iv 2, where Hume speaks of the “source of the error and deception with regards to identity, when we attribute it to our resembling perceptions, notwithstanding their interruption” (T 202). Hume’s T 255 reference to “identity, in an improper sense” perhaps expresses some inclination to classify the error as verbal, but, on the whole, he seems convinced that it is more than verbal. It is, after all, at odds with the metaphysical truth that every distinct perception is a distinct existence, no less than is (iii) or (iv).

The difference between tendency (i) and the others must therefore lie elsewhere. In my view it comes to this: (i) is a deeper or more fundamental tendency, and its depth has to do (though only in part) with its psychological indispensability. As Hume says at T 254, “we cannot long sustain” resistance to (i). Tendency (ii) is also indispensable, but it is less fundamental: it owes its

influence to the earlier work of (i), and to our contradiction-sensitivity. Tendencies (iii) and (iv) are, like (ii), also triggered by (i), but unlike (ii), they are dispensable. Tendencies (iii) and (iv) go into operation only when we attempt "to justify to ourselves [the] absurdity" produced by (i). The same is true of (ii), but (ii) is an indispensable tendency because we need it to avoid the most extreme form of unjustification: contradiction. The unjustification threatening those who abstain from (iii) and (iv) is much more mild. Hume acknowledges that there are times when (iii) and (iv) do not take effect. At those times, he explains, "we still feel a propensity" to confound ideas, as (i) demands, "tho' we are not able fully to satisfy ourselves in that particular, nor find any thing invariable and uninterrupted to justify our notion of identity" (T 255).<sup>45</sup>

The lesson I draw from all this is not that tendency (i) lies beyond lasting philosophical criticism—it is, as we saw, inconsistent with the metaphysical principle that distinct perceptions are distinct existences—but that it is markedly less subject to criticism than (iii) and (iv). This is partly because (i) is indispensable, and partly because the fictions produced by (iii) and (iv) occasion *further* conflict with the metaphysical principle—conflict that can be avoided.

Now, tendency (i) does not as it stands play a role in Hume's account of *personal* identity. And it cannot do so, because persons as Hume understands them are not single perceptions, but heaps or bundles of perceptions. So at T 260 and T 261, after turning from bodies to persons, Hume puts two tendencies in place of (i): the first, (i\*), a tendency to suppose that resembling perceptions—or collections of resembling perceptions—constitute one continued object; and the second, (i\*\*), a tendency to think that causally related perceptions—or collections of causally related perceptions—constitute one continued object.

Tendency (i\*) does raise difficulties, but Hume's troubles in the Appendix have to do, I believe, with (i\*\*). Hume gives many indications, in both I iv 6 and the Appendix, that resemblance—the relation that triggers (i\*)—plays a relatively minor role in accounting for our belief in personal identity and simplicity.

First, what Hume calls "the true idea of the human mind" involves causation rather than resemblance. "The true idea of the human mind"—the idea that reveals why every one of the mind's perceptions is included there—"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" (T 261).<sup>46</sup>

Second, Hume says explicitly that memory is "chiefly" the source of personal identity not insofar as it generates resemblance, but insofar as it supplies us with the notion of cause and effect, and with the knowledge of a particular system of causally related perceptions (T 261).

Third, only cause and effect can explain how we extend identity beyond memory, in order to “comprehend times, and circumstances, and actions, which we have entirely forgot” (T 262).

Fourth, identity with regard to the passions corroborates the importance of causation, and *only* the importance of causation. At T 253, when he introduces the topic of personal identity, Hume, as we saw, warns that “we must distinguish betwixt personal identity, as it regards our thought or imagination, and as it regards our passions or the concern we take in ourselves.” One reason for the warning is that identity as it regards our passions would not corroborate what Hume says in I iv 6 about resemblance. When, after describing the true idea of the human mind, Hume writes that “*in this view* our identity with regard to the passions serves to corroborate that with regard to the imagination” (T 261, my emphasis), the “view” he has in mind is the view of the mind as a system of causally related perceptions. Attributions of identity to the understanding, which includes imagination and memory, do, up to a point, involve resemblance. But when Hume’s account is extended into an account of the whole person, as in Book II, “Of the Passions,” the relation that matters, or matters primarily, is not resemblance but cause and effect, as Jane L. McIntyre has shown.<sup>47</sup>

Finally, the two principles Hume identifies in the climactic passage of the Appendix at T 636—the distinct existence of distinct perceptions, and the mind’s failure to perceive real connections among distinct existences—though not wholly irrelevant to resemblance, bear more particularly on causation, which is, according to Hume, commonly understood as a real connection.

What, then, is the problem with (i\*\*)? Why can’t it be said that we treat causally related objects as one thing? My reply is that causal relations are, in Hume’s view, just not enough, and to explain why, I want to enlist the help of Shaftesbury. As we saw earlier, Hume portrays Shaftesbury as an ingenious but unfortunate victim of trivial propensities of the imagination—tendencies (iii) and (iv). A reading of Shaftesbury’s *Moralists* confirms Hume’s report. The inferences Theocles defends there have the following general structure: they begin with the observation that something is one (or with the observation that its parts stand in the kind of causal relations required to make them one), and then infer, usually as an explanation of its observed oneness, a hidden (or as Hume would say, “unknown and mysterious”) nature or principle of union. We saw such an inference in the case of a tree; Theocles also makes it in the case of the whole universe. But Theocles, as we also saw, makes what can be called a parity objection to Philocles: how can you affirm your own oneness, and yet deny the oneness of the universe? Theocles is raising, I believe, a question of justification, but in asking it he could have raised, in Hume’s mind, a question of explanation. How do I explain my thinking that I am one, while I am nonetheless certain that . . . nature is not? This question brings home the limitations of (i\*\*). How can causal relations among my parts explain why I take

myself to be one—a self that enjoys “real simplicity and identity,” as Hume puts it at T 635, or a self that is invariable and uninterrupted, as he puts it elsewhere—when the same relations do not lead me to think of nature or the universe as one? There are, of course, some replies open to Hume. He can point out that my perceptions are causally related in a special way: they sympathize to a common end. But this could hardly impress a close reader of Shaftesbury, who (not at all implausibly) finds common ends, and parts cooperating to achieve them, throughout nature. Hume might instead bite the bullet—it’s not such a hard bullet, after all—and say that all of us, himself included, think of nature or the universe as one.<sup>48</sup> But Shaftesbury finds the relevant causal relations in too many places. As Theocles says,

All things in this World are *united*. For as the *Branch* is united with the *Tree*, so is the *Tree* as immediately with the *Earth*, *Air*, and *Water*, which feed it. As much as the fertile *Mould* is fitted to the *Tree*, as much as the strong and upright *Trunk* of the *Oak* or *Elm* is fitted to the twining *Branches* of the *Vine* or *Ivy*; so much are the very *Leaves*, the *Seeds*, and *Fruits* of these *Trees* fitted to the various *Animals*: These again to one another, and to the *Elements* where they live, and to which they are, as *Appendices*, in a manner fitted and join’d; as either by *Wings* for the *Air*, *Fins* for the *Waters*, *Feet* for the *Earth*, and by other correspondent inward *Parts* of a more curious *Frame* and *Texture*. (M 2: 287; C 274)

This is enthusiasm, as Philocles points out and even Theocles admits. But it presents a problem for Hume. Note that Theocles isn’t defending the conclusion of his characteristic inferential structure, but his premise. He is concerned here not with (iii) or (iv), but with something like (i\*\*). If I think of myself as one because of the kind of causal relations Hume follows Shaftesbury in pointing to, why don’t I think of tree-and-vine as one, or of cone-and-crossbill, or of mulch-and-worm? Hume tells us in the Appendix that causation cannot unite successive perceptions in our thought or consciousness. I think these Shaftesburian texts point to the reason why. Causation cannot explain why we unite the perceptions we call a mind or person because it is equally present when we experience little or no inclination to unify, much less an inclination to think of the unified complex as simple and strictly identical.<sup>49</sup>

If my proposal is correct, it should allow me to explain why, in the Appendix, Hume is not unhappy with his account of our belief in bodily identity, even though the causal relations too weak to account for our belief in personal identity also play a role there. The explanation is simple: attributions of bodily identity are more dependent on resemblance: bodies are not, at least in appearance, “all in revolution” as minds are.<sup>50</sup>

My proposal also allows me to explain why Hume did not lose confidence in the power of causation to account completely for some kinds of perceived

unity, even though he had concluded that it was too weak to account for the peculiar unity we ascribe to minds or persons. Consider in this connection a little-known passage from Section III of Hume's first *Enquiry*—little-known because it was part of Section III only between 1748 and 1772, before Hume, preparing for what became the posthumous edition of 1777, apparently asked that it be cut. (It is therefore omitted from P. H. Nidditch's edition.)<sup>51</sup> In the excised passage, Hume uses contiguity, resemblance, and causation to explain "a kind of *Unity*"—the unity of a "composition of genius" or literary production. "In narrative compositions," he writes, "the events or actions, which the writer relates, must be connected together, by some bond or tye: They must be related to each other in the imagination, and form a kind of *Unity*, which may bring them under one plan or view, and which may be the object or end of the writer in his first undertaking" (Beauchamp, EHU 102). Elsewhere in the excised pages he speaks of "a species of unity" (Beauchamp, EHU 103) or "a certain unity" (Beauchamp, EHU 103, 106). The associative principle responsible for the unity varies. "An annalist or historian who should undertake to write the history of EUROPE during any century," for example, "would be influenced by the connexion of contiguity in time and place." But "the most usual species of connexion among the different events, which enter into any narrative composition," Hume explains, "is that of cause and effect; while the historian traces the series of actions according to their natural order, remounts to their secret springs and principles, and delineates their most remote consequences" (Beauchamp, EHU 103). The historian understands that "the more unbroken the chain is, which he presents to his reader, the more perfect is his production." Knowledge of causes "is not only the most satisfactory," causation being "the strongest of all" relations or connections; it is also "the most instructive; since it is by this knowledge alone, we are enabled to controul events, and govern futurity" (Beauchamp, EHU 103). Later he says that the relation of cause and effect "unites [several events] into one body" (Beauchamp, EHU 106).

This passage has a bearing in the present context because it is a positive, post-Appendix use of principles of association—notably causation—to explain the same sort of psychological operation Hume had hoped to account for in *Treatise* I iv 6. The passage makes it clear that after the Appendix, Hume did not lose faith in the ability of causation to explain at least one kind of perceived unity: the unity we find in a well-constructed historical narrative, or—to bring the passage closer to the topic of personal identity—in a well-constructed work of biography. If the same principle cannot explain why we take a person to be a unity, perhaps the reason is simply that the unity we find in a biography, however tightly it is woven, always falls short of the unity we attribute to the person whose life it chronicles.<sup>52</sup>

Earlier I suggested that Hume's unhappiness in the Appendix did not extend to bodies, because over time (at least when judged on a human scale),

bodies change less dramatically than minds do. But Hume's confidence in his treatment of body may have had a second source. At least in *Treatise* I iv 6, something led Hume to take the boundaries of bodies—their unity at a time—for granted. Perhaps he thought that the spatial contiguity of their parts entitled him to do this. (See, in this connection, T 255, where we are asked to "suppose any mass of matter, of which the parts are contiguous and connected," to be placed before us.) Whatever the reason, his problem was understanding why we take a body with established boundaries to be the same over time; he didn't ask how the boundaries are established in the first place. In the case of minds, however, Hume is explicitly concerned with the establishment of boundaries at a time, and it is here, where the concern is unity at a time ("simplicity") rather than identity over time, that the insufficiency of causal relations is perhaps most apparent.

Hume's account of our belief in our simplicity is offered, without elaboration, on the concluding page of *Treatise* I iv 6, as if Hume took it to be merely supplemental:<sup>53</sup>

What I have said concerning the first origin and uncertainty of our notion of identity, as apply'd to the human mind, may be extended with little or no variation to that of *simplicity*. An object, whose different co-existent parts are bound together by a close relation, operates upon the imagination after much the same manner as one perfectly simple and indivisible, and requires not a much greater stretch of thought in order to its conception. From this similarity of operation we attribute a simplicity to it, and feign a principle of union as the support of this simplicity, and the center of all the different parts and qualities of the object. (T 263)

The account is formulated so that it applies to bodies as well as minds, but it remains true that Hume was preoccupied more with the simplicity of mind than the simplicity of body. This may be because he understood simplicity as "uncompoundedness" or "indivisibility" (as is suggested not only by T 263 itself, but by Hume's remarks on the simplicity of soul and Spinozist substance in *Treatise* I iv 5), and believed that the human tendency to attribute uncompoundedness or indivisibility was stronger in the case of mind than in the case of body.

Whatever the explanation, there is an obvious obstacle to an account of simplicity that turns on cause and effect. According to Hume, causation calls for temporal priority or *non*-coexistence (see T 76, as well as T 31). How then can causation be the relation that unites, in imagination, the "different co-existent parts" (T 263) of the mind?

It might be thought that I should take these questions, and the challenge they raise, as support for my interpretation. I am, however, unwilling to do so, partly because there can be *indirect* causal relations even among co-existent

objects,<sup>54</sup> and partly because Hume does not always understand “co-existent” in a strict or narrow way. On the strict or narrow understanding, two objects co-exist only if they occupy the same indivisible moment of time. But at T 237, Hume seems to rely on a broader understanding of co-existence, one that allows cause-and-effect relations among “co-existent” objects. On the broad understanding at work there, objects seem to count as co-existent so long as the moments of time they occupy are very close. The same might be said of the “co-existent parts” at T 263. They may be “co-existent” because they exist at *around* the same time, in which case there will be room for causal relations between them.<sup>55</sup> The challenge raised by my earlier questions can therefore be met. But I do think Hume came to realize that causal relations cannot account for judgments of simplicity, because causal relations of comparable strength are found, as Shaftesbury’s characters suggest, in far too many circumstances—circumstances in which we feel no inclination to make either judgments of simplicity or weaker judgments of unity.

I have argued that Shaftesbury’s *Moralists* points to a problem with Hume’s (i\*\*), the crucial element in his account of our belief in personal identity, and that Hume came, in the Appendix, to see the problem himself. I want to offer one last piece of support: the fact that tendency (i\*\*)—the tendency to think that causally related perceptions, or collections of causally related perceptions, constitute one continued object—is not as easy to hold apart from tendencies (iii) and (iv) as (i) was.

We are able to hold (i) apart from (iii) because in (i), we suppose merely that objects or perceptions continue when they are unperceived, while in (iii), we imagine that objects or perceptions enjoy the support of an enduring soul, self, or substance. (For simplicity’s sake, I will confine my remarks to the relationship between (i\*\*) and (iii).) In the case of (i\*\*), we think that causally related objects constitute one object, without thinking—if (i\*\*) is to be distinct from (iii)—that they exist together in a self or substance. But then what, exactly, is the thought produced by (i\*\*)? How, exactly, does it involve a belief in “real” or “perfect” simplicity and identity (T 635)? (Note that (i\*\*) cannot really be clarified by (i), which applies most convincingly to single objects or perceptions that are very much alike.)

A neglected feature of T 254 will, I hope, lend force to my claim that (i\*\*) is more problematic than (i). When Hume writes that “we feign the continu’d existence of the perceptions of our senses, to remove the interruption; and we run into the notion of a *soul*, and *self*, and *substance*, to disguise the variation,” he forges two associations: the first between the fiction of continued existence and the removal of interruption; the second between the notion of soul, self, or substance and the disguise of variation. The “disguise of variation” is exactly what (i\*\*) needs to achieve, and I think it is noteworthy that here it is linked with a *dispensable* tendency. The question is: is there an indispensable tendency that can bring about the same effect?

At T 215–216, Hume writes that were we fully convinced that perceptions are dependent, interrupted, and different, “we shou’d clearly perceive the error of our first supposition of a continu’d existence, and wou’d never regard it any farther.” The “first” supposition of continued existence is, I take it, the vulgar supposition of continued existence, the upshot of tendencies (i) and (ii) in my original list of four. Hume is telling us that if we could see (i) or (ii) as mistaken, we would not make the supposition. But we cannot see them as mistaken, at least not for long, because the tendencies in (i) and (ii) are too deeply rooted. I am proposing that although full conviction isn’t possible here, we can become fully convinced of the distinctness of (at least most of) the perceptions thought to constitute the mind. Hume doesn’t see any version of (i\*\*) that can explain why anyone, even an ordinary person, comes to think of many objects as constituting one object merely because they are causally related. Mere causal relations just aren’t enough, especially for the philosopher, who is fully and stably persuaded that the many objects in question are distinct.

I have suggested that Hume thinks he cannot explain why we think of ourselves as one thing over time, or as strongly unified at a time. His worry, I think, is why *anyone* would think of himself or herself in this way, but as I have already indicated, philosophers may present him with a more intense version of the difficulty.<sup>56</sup> Sections 5 and 6 of *Treatise* I iv are the climax of a long development, in which readers are persuaded that there are no real connections (or at least no perceived real connections) among perceptions. If philosophers can maintain their commitment to this metaphysics, it isn’t easy to see why they should believe in their own identity. Felt connections, once identified and properly classified, merely enlarge the mind’s population, and have no power to unite it.<sup>57</sup> But philosophers, at least Humean philosophers, cannot abandon belief in their own identity, because their identity is implicated in the ethical concern that is the stimulus for their scientific project.<sup>58</sup>

I cannot forbear having a curiosity to be acquainted with the principles of moral good and evil, the nature and foundation of government, and the cause of those several passions and inclinations, which actuate and govern me. I am uneasy to think I approve of one object, and disapprove of another; call one thing beautiful, and another deform’d; decide concerning truth and falshood, reason and folly, without knowing upon what principles I proceed. (T 270–271).

It might be said that a reductionist or relational analysis of discourse about the self can allow for all this. But Hume needs more than an analysis of the sentences I’ve quoted. He needs an agent of unification—a force that will explain why he believes in his own identity, and why he takes it to be as tight as he does—and this he does not seem to have.

If I am right, then, in the Appendix Hume faces an explanatory failure, one that owes its weight to the ethical concern behind it, and its general shape

to a metaphysics that genuinely improves upon his predecessors. Hume's metaphysical achievements should not be underestimated. He explains, by the appeal to causation, how a later person can be the same as an earlier one even when the earlier person's thoughts and deeds have been entirely forgotten. This improves on Locke and lends supports (at least up to a point) to Shaftesbury. But Hume also explains how a person can be the same even after dramatic change (or revolution) in character or disposition, even if the person is in no way simple (T 261). This corrects Shaftesbury. Hume explains, in effect, why our identity consists no more in the will than it does in the understanding, and he is able to explain why philosophers are sometimes drawn to the image of the self as strangely simple. Our identity is a matter of causation rather than persistence. If it *appears* to lie in the persistence of something simple, that is because our identity is complex in a way we were not predisposed to recognize.

Shaftesbury renounces metaphysics (or the full elaboration of a metaphysics), sometimes in favor of ethics, sometimes in favor of a richly descriptive natural history. Hume reinstates metaphysics, but his philosophizing has an ethical aim that is very close to Shaftesbury's: he wants to achieve a kind of unity, consistency, or integrity. He wants to know the principles upon which he proceeds, and he wants to be able to approve of them. But this very aim presupposes a commitment that his metaphysics leaves him powerless to explain. In the Appendix, Hume observes that were the true metaphysics of the person different, he could explain why we believe in our own identity. If perceptions had to inhere in something simple and individual, or if there were real connections between them, "there wou'd be no difficulty in the case" (T 636). But the true metaphysics of the person is what it is, and it is, philosophically, wholly satisfying. There is no philosophical objection to it.

I would like to close with a brief consideration of two objections.

My account resembles one that Don Garrett attributes to A. H. Basson and S. C. Patten, according to which (in Garrett's words) "causation and resemblance ... are simply not, as a matter of fact, adequate to the scope of the job" assigned to them in *Treatise* I iv 6 (T 172).<sup>59</sup> This account is also discussed, as Garrett observes, by Barry Stroud, who dismisses it with the following words:

Does Hume recognize that, as I have suggested, the natural relations of resemblance and causality are not actually enough to do the job required of them, given the nature of our experience? It seems unlikely; that could easily have been said, and illustrated.<sup>60</sup>

But my claim is that it is causation in particular that cannot do the job assigned to it, and it seems to me that although Hume does not say so explicitly, he gives several indications that this is his view. I admit that Hume could easily have said what I have said. Stroud's point, I take it, is that there is a

presumption against explanations as simple as the one I have provided. But surely there is an equally weighty presumption against the complex explanations so common in the literature: precisely because they are not easy to state, they are more distant from the actual text of the Appendix, which Hume apparently regarded as an adequate statement of his difficulty.<sup>61</sup>

But what about his talk of contradiction? The two principles identified at T 636 are certainly consistent. But Hume is unable to see how they can both be true. If either one were false, as he goes on to say, there would be no problem in the case. This isn't quite what we would call a contradiction or inconsistency, but neither is the following, from T 268 in I iv 7:

Very refin'd reflections have little or no influence upon us; and yet we do not, and cannot establish it for a rule, that they ought not to have any influence; which implies a manifest contradiction.

Hume's quick resort to talk of contradiction may not be strictly speaking justified, but it is certainly understandable. If his two principles are true, Hume cannot explain why we believe in our identity and unity. And yet we do. Hume cannot claim, then, to offer us all of the following: a metaphysically adequate account of persons; a complete list of associative principles; and a successful science of human nature—one that explains everything it should, among them the ethical concern of which it is so admirable an expression.

## NOTES

Earlier versions of this paper were delivered at the 26th International Hume Society Conference at University College, Cork, in July 1999, to the Department of Philosophy at the University of Virginia in November 1999, and at the Central Division Meeting of the American Philosophical Association in April 2000. I am grateful to audience members (particularly Don Garrett, Harold Langsam, Wade Robison, and Galen Strawson) for their questions, and to Donald Baxter, Michael Lynch, and Udo Thiel for their written comments. I also owe thanks to Aaron Garrett, Knud Haakonssen, and the members of their Locke seminar at Boston University, for the opportunity to discuss, in March 2000, some of this material with them. I am particularly indebted to Jane McIntyre, my commentator at the Central Division Meeting, for thoughtful criticisms that led to several changes.

1. Quotations from Locke's *Essay*, identified in parentheses by book, chapter, and section number, are taken from John Locke, *An Essay concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Clarendon Press, 1975).

2. Recent writing on personal identity raises the same question, even when ethical considerations are not the focus of concern. For example, at one point in his very stimulating book *The Human Animal* (New York: Oxford University Press, 1997),

Eric T. Olson distinguishes between my identity in what he calls the “practical sense” and my identity in the “numerical sense” (69). Whether I should be held responsible for an earlier misdeed turns, he thinks, on my identity in the first sense; it is a question to be settled by moralists rather than by metaphysicians. Olson’s own concern centers on my identity in the second sense—a matter for metaphysics rather than morals. Olson preserves the independence of his metaphysical claims by disconnecting them from Locke’s forensic preoccupations. See also Carol Rovane, *The Bounds of Agency: An Essay in Revisionary Metaphysics* (Princeton: Princeton University Press, 1998), 69.

3. Quotations from David Hume’s *Treatise* are identified in parentheses by “T,” followed by page number in David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2nd ed. revised by P. H. Nidditch (Oxford: Clarendon Press, 1978), from which I quote throughout.

4. Locke’s account of personal identity is more “ethical” than Shaftesbury supposes, but for the most part, I ignore this complication here. I defend a more ethical reading of Locke in “Locke on Personal Identity,” *Journal of the History of Philosophy* 29 (1991): 201–226, reprinted in Vere Chappell (ed.), *Locke* (Oxford: Oxford University Press, 1998), 149–174. Other more ethical readings of Locke—readings emphasizing concern, appropriation, or self-constitution—include David P. Behan, “Locke on Persons and Personal Identity,” *Canadian Journal of Philosophy* 9 (1979): 53–75; Udo Thiel, *Lockes Theorie der personalen Identität* (Bonn: Bouvier, 1983); Marya Schechtman, *The Constitution of Selves* (Ithaca, N.Y.: Cornell University Press, 1996), 105–112; Thiel, “Personal Identity,” in Daniel Garber and Michael Ayers (eds.), *The Cambridge History of Seventeenth-Century Philosophy* (Cambridge: Cambridge University Press, 1998), vol. 1, 888–899; and Raymond Martin and John Barresi, *Naturalization of the Soul: Self and Personal Identity in the Eighteenth Century* (London: Routledge, 2000), 18–29. See also Raymond Martin, *Self-Concern: An Experiential Approach to What Matters in Survival* (Cambridge: Cambridge University Press, 1998), 4 and 107; Olson, *The Human Animal*, 57–8; and Carol Rovane, *The Bounds of Agency*, 45, 70, 234. For reservations about such readings see Michael Ayers, *Locke* (London: Routledge, 1991), vol. 2, 266–268.

5. Passages from the *Characteristicks* are cited and quoted as they appear in Anthony Ashley Cooper, Third Earl of Shaftesbury, *Characteristicks of Men, Manners, Opinions, Times* (London, 1711; reprinted Hildesheim: Georg Olms, 1978), 3 vols. Throughout the paper I provide, in parentheses, an abbreviated title of the particular work cited, followed by volume and page number in this collection. Each citation also includes a reference to the passage as it appears in *Characteristicks of Men, Manners, Opinions, Times*, ed. Lawrence E. Klein (Cambridge: Cambridge University Press, 1999). Klein’s text is based on the second (1714) edition of the *Characteristicks*. I use the following abbreviations:

- A: *Soliloquy, or Advice to an Author* (appearing in vol. 1 of the 1711 edition);
- I: *An Inquiry concerning Virtue or Merit* (in vol. 2);
- M: *The Moralists: A Philosophical Rhapsody* (also in vol. 2);
- MR: *Miscellaneous Reflections on the said Treatises, and other Critical Subjects* (occupying the whole of vol. 3);
- C: Klein’s edition of the *Characteristicks*.

For discussions of Shaftesbury on personal identity, and of the influence of his treatment on Hume’s, see Ben Mijuskovic, “Hume and Shaftesbury on the Self,”

*Philosophical Quarterly* 21 (1971): 324–337; Clive M. Corcoran, "Do We Have a Shaftesburean Self in the *Treatise*?" *Philosophical Quarterly* 23 (1973): 67–72; Ben Lazare Mijuskovic, *The Achilles of Rationalist Arguments* (The Hague: Martinus Nijhoff, 1974), 103–114; and Martin and Barresi, *Naturalization of the Soul*, 60–64 and 183. There are brief but illuminating remarks on Shaftesbury's view of the self in Stephen Darwall, *The British Moralists and the Internal "Ought": 1640–1740* (Cambridge: Cambridge University Press, 1995), 180 and 197–204, and in J. B. Schneewind, *The Invention of Autonomy: A History of Modern Moral Philosophy* (Cambridge: Cambridge University Press, 1998), 307–308. Michael Gill discusses some closely related issues in "Shaftesbury's Two Accounts of the Reason to be Virtuous," a paper delivered at the 26th International Hume Society Conference, Cork, Ireland, 1999, and forthcoming in the *Journal of the History of Philosophy*. For general accounts of Shaftesbury's moral philosophy see Darwall, *The British Moralists*, 176–206, Schneewind, *The Invention of Autonomy*, 295–309, and Charles Taylor, *Sources of the Self: The Making of Modern Identity* (Cambridge, Mass.: Harvard University Press, 1989), 248–59. Shaftesbury's larger political and cultural context is helpfully described by Klein in the introduction to his edition, vii–xxxii.

6. See also A 1: 291 and C 130, criticizing "these Physiologists and Searchers of Modes and Substances."

7. For Taylor's account of strong evaluation, see, for example, "What Is Human Agency?" 15–44, and "The Concept of a Person," 97–114, in *Human Agency and Language*, vol. 1 of his *Philosophical Papers* (Cambridge: Cambridge University Press, 1985).

8. I wish I had space to discuss all of the ways in which Hume's reading of Shaftesbury helped to shape his literary consciousness. It was Shaftesbury's rich description of his own inner life that led Hume to count him, despite his many "unscientific" qualities as a thinker and a writer, as an important figure in the emerging science of man.

9. I agree with both Annette C. Baier and Christine M. Korsgaard that Hume is seeking "reflexive self-acceptance" or "reflective endorsement." See Baier, *A Progress of Sentiments* (Cambridge, Mass.: Harvard University Press, 1991), and Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), 51–66. Shaftesbury is seeking it as well. Hume, in a passage emphasized by Baier, speaks of the virtuous mind's ability to "bear its own survey" (T 620). And Shaftesbury describes the virtuous mind as "*such as can freely bear its own Inspection and Review*" (I 2: 114; C 206).

10. Martin and Barresi survey some of Shaftesbury's objections to Locke in their *Naturalization of the Soul*, 62–63.

11. See *The Works of George Berkeley*, ed. A. A. Luce and T. E. Jessop (London: Thomas Nelson, 1948–57), vol. 3, 315.

12. For an account of Hume's response to excessive Pyrrhonism, see my paper "Hume's Inductive Skepticism," in Margaret Atherton (ed.), *The Empiricists: Critical Essays on Locke, Berkeley, and Hume* (Lanham, Md.: Rowman and Littlefield, 1999), 183–212.

13. Martin and Barresi also take note of this. See *Naturalization of the Soul*, 60–61.

14. *The Life, Unpublished Letters, and Philosophical Regimen of Anthony, Earl of Shaftesbury*, ed. Benjamin Rand (London: Swan Sonnenschein, 1900), 136.
15. *Ibid.*
16. *Ibid.*, 138. Note the very clever concluding sentence. Shaftesbury and Pyrrho have their “individual understandings”—that is, their individual opinions. But they could not have them, Shaftesbury is saying, without being “of two minds” in another sense—without having distinct individual minds.
17. Locke’s analogy between life and consciousness—“a great analogy,” said Hume (T 253), who followed Locke in stressing it—helps to insulate our practice from our ignorance, as do his many thought experiments, which are, for the most part, designed to show that the leading conjectures about the underlying reality—materialism and Cartesianism—are irrelevant to our concerns (or that the competition between them is irrelevant). On the latter point, see Locke’s defense, in § 27, against the potential “absurdity”—by which I think he means the real impossibility—of some of his suppositions.
18. As Schneewind writes, though in a different context, Shaftesbury neutralizes skepticism “by accepting that even if all we have, in morality, are appearances, the appearances suffice” (*The Invention of Autonomy*, 305).
19. *The British Moralists*, 180.
20. What I here call the “classical problem” is what Marya Schechtman calls the “reidentification question,” the question of what it is that makes someone the same person at different times (*The Constitution of Selves*, 2). Shaftesbury is sometimes addressing not this problem, but what Schechtman calls the “characterization question,” the question of what characteristics make someone the person he or she is (74). Schechtman argues (for example, on 77–78) that her two questions are distinct though importantly related.
21. For Parfit’s understanding of “what matters,” see in particular ch. 12 of *Reasons and Persons* (Oxford: Clarendon Press, 1984), 245–306.
22. I borrow the word “punctual” from *Sources of the Self*, 159–174, where Charles Taylor argues (incorrectly, in my view) that Locke’s self is punctual.
23. Survival in such a case is possible, Hume thinks, because “whatever changes he endures,” the person’s parts remain connected by causation (T 261).
24. Jane L. McIntyre takes Shaftesbury to accept the self’s strange simplicity; see “Personal Identity and the Passions,” *Journal of the History of Philosophy* 27 (1989): 548. Martin and Barresi agree; see *Naturalization of the Soul*, 62, where they observe that it is unclear how Shaftesbury would account for our strange simplicity. See also Mijuskovic, *The Achilles of Rationalist Arguments*, 103–105. Hume may have been the first of Shaftesbury’s readers to suppose that he meant to endorse the self’s “strange simplicity.” Hume doesn’t say so explicitly, but he may think that the simplicity of a Shaftesburian inner principle is like that of the immaterial soul of the theologians. That soul issues in various modifications, without itself undergoing variation (T 241). Hume takes this to be unintelligible or absurd, and he may think the same of Shaftesbury’s simple self.
25. In *Philosophical Dialogue in the British Enlightenment* (Cambridge: Cambridge University Press, 1996), 47–73, Michael Prince argues that Theocles and Philocles

represent "two opposing tendencies" in Shaftesbury (69). "In all likelihood," Prince suggests, "we are meant to read *The Moralists* both as a heroic drama leading to the victory of Theocles and as a letter indicating the skeptic's ongoing resistance to any such conversion. The generic structure moves the reader from a metaphysical either/or to an ironic both/and" (69).

26. Darwall observes that Shaftesbury champions a "rational" enthusiasm (see *The British Moralists*, 189), but I am not persuaded that Shaftesbury *wholeheartedly* embraces the enthusiasm of Theocles. (Schneewind's reading of Shaftesbury's response to enthusiasm differs strikingly from Darwall's; see *The Invention of Autonomy*, 303, particularly n. 31. For useful remarks on the ambiguity in Shaftesbury's understanding of enthusiasm, see Lawrence E. Klein, *Shaftesbury and the Culture of Politeness* [Cambridge: Cambridge University Press, 1994], 165–168 and 172–173.) It must, however, be admitted that the author of the *Miscellaneous Reflections* sides with Theocles rather than Philocles—with the "hero" of the dialogue rather than its "skeptic." For Shaftesbury's contrast between the dialogical "hero" and his skeptical adversary, see MR 3: 295 and C 463. Shaftesbury's contrast helps to clarify Hume's observation, in a 1751 letter to Gilbert Elliot of Minto, that Cleanthes is the hero of the *Dialogues concerning Natural Religion* (*The Letters of David Hume*, ed. J. Y. T. Greig [Oxford: Clarendon Press, 1932], vol. 1, 153). Hume's observation is not an expression of sympathy for Cleanthes, but a sign of allegiance to a literary form (which is not to say that Hume is unwilling to have Elliot understand him in another way). Hume goes on to assure Elliot that he is open to any suggestion that will strengthen the hero's side of the argument. Here too Hume may be following Shaftesbury, who advises the writer of philosophical dialogues on religious themes to make the hero "as dazzling *bright*, as you are able" (MR 3: 295; C 463). (Shaftesbury also advises the writer to "be not so cautious of furnishing your *Representative SCEPTICK* with too good *Arguments*, or too shrewd a Turn of *Wit* or *Humour*." Hume clearly followed this advice. "If when you have fairly wrought up your *Antagonist* to his due Strength and cognizable Proportion," Shaftesbury continues, "your *chief Character* cannot afterwards prove a match for him, or shine with a superiour Brightness," the fault lies not with the subject—"this, I hope, you will never allow"—but with the author.)

27. This, of course, leads to the question, what makes the perfected and to-be-perfected selves one and the same? The only available answer, it may seem, is their common possession of an atom of selfhood, in which case even the perfected self is in danger of becoming strangely simple. I myself am uncertain that this is the only answer available to Shaftesbury, but because I do not think he accepts the interpretation of strange simplicity embodied in the question, I will leave the matter here.

28. If this is correct, then the self is unlike an atom, despite Theocles's comparison between them (M 2: 352). That Theocles makes the comparison does weigh against the interpretation I have presented here, but there may be no single interpretation consistent with every detail of this highly polyphonic text. Its many voices include not only the author of *The Moralists*, but Philocles the narrator (who is reporting his conversation with Theocles in a letter to Palemon), Philocles as he appears within the narration, Theocles, and (in volume 3 of the *Characteristicks*) the miscellaneous author who reflects on *The Moralists*. For readings of Shaftesbury attentive to these complexities see Klein, *Shaftesbury and the Culture of Politeness*, 113–119, and Prince, *Philosophical Dialogue*, 23–73.

29. It might be asked whether there is *any* sense in which these governing principles are simple, if their simplicity is not to be understood as absolute unanalyzability. Shaftesbury is, I am afraid, not very helpful here. But the claim of simplicity means at least this: the inner principles that bear my identity are not material things; and their unity (unlike the observed integrity of the human organism they animate) has no further ground.

30. That Hume was sensitive to such matters is made clear by the very Lockean title he originally gave to the first *Enquiry: Philosophical Essays concerning Human Understanding*. For further discussion, see the editor's introduction to my abridgment of John Locke, *An Essay concerning Human Understanding* (Indianapolis: Hackett, 1996), xvi–xviii. Hume responds to the meaning of “essay” at *An Enquiry concerning Human Understanding*, ed. Tom L. Beauchamp (Oxford: Oxford University Press, 1999), 102. (Subsequent references to this edition are in the form “Beauchamp, EHU,” followed by page number.) Shaftesbury comments on its meaning in several places, for example at M 2: 265; C 266 and MR 3: 96–97; C 380. Hume seems to be reworking the imagery of the latter passage at T 263–264.

31. The paradigmatic objects of Lockean consciousness are, perhaps, thoughts or ideas, but as this passage indicates, those objects include the body and its parts, as it after all must if “actions,” as well as thoughts, are to fall within it. In this connection it is interesting to consider T 191, where Hume writes that “properly speaking, 'tis not our body we perceive, when we regard our limbs and members, but certain impressions, which enter by the senses.”

32. For further evidence of the body's restoration in Book II, see Gordon Park Stevenson, “Humean Self-Consciousness Explained,” *Hume Studies* 24 (1998): 120–121, and Jane McIntyre, “Hume's Underground Self,” *Studies in Early Modern Philosophy* 3 (1993): 110–126. I think it is very likely that Hume has the body at least partly in mind when he writes, at T 313, of the “idea, or rather impression of ourselves” that is “always intimately present with us.” See Ayers, *Locke*, vol. 2, 285–287, for a compelling account of the way in which awareness of ourselves as material objects permeates all of our experience.

33. For Parfit's statement of reductionism see his *Reasons and Persons*, 210; for Martin on relational accounts of personal identity see his *Self-Concern*, 1.

34. See, for example, Phillip D. Cummins, “Hume's Disavowal of the *Treatise*,” *Philosophical Review* 82 (1973): 379; Barry Stroud, *Hume* (London: Routledge, 1977), 133–4; Wade L. Robison, “In Defense of Hume's *Appendix*,” in David Fate Norton, Nicholas Capaldi, and Wade L. Robison (eds.), *McGill Hume Studies* (San Diego: Austin Hill Press, 1979), 96; Don Garrett, *Cognition and Commitment in Hume's Philosophy* (New York: Oxford, 1996), 179; Donald L. M. Baxter, “Hume's Labyrinth Concerning the Idea of Personal Identity,” *Hume Studies* 24 (1998): 203–233; and Donald C. Ainslie, “Hume's Reflections on the Identity and Simplicity of Mind,” *Philosophy and Phenomenological Research*, forthcoming.

35. There is one potential sticking point. The quoted sentence says only that we have no notion of the mind distinct from particular perceptions. It does not say that the mind is nothing but particular perceptions. At least in the present case, Hume leaves room for what might be called a “skeptical realist” account of personal identity, according to which there is (or may be) a mind or self distinct from perceptions, even if we have (or can have) no contentful notion of it.

36. A noteworthy feature of the passage is Hume's description of the evidence of his metaphysics (or epistemology) as "seeming." This does insinuate the possibility that this evidence—"evidentness," as Don Garrett and David Owen call it—is misleading. (On this meaning of "evidence," see Garrett, *Cognition and Commitment in Hume's Philosophy*, 228; Owen, *Hume's Reason* (Oxford: Oxford University Press, 1999), 185–188; and David Hume, *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton [Oxford: Oxford University Press, 2000], 575.) I don't want to rule this out. When I say that Hume regards his metaphysics as successful, I do not mean he is a dogmatist. If the proposal I go on to develop is correct, his metaphysics leaves him with a problem, and since he claims to be open to solutions (T 636), in the end he may have to face the possibility that his metaphysics is after all mistaken.

37. In the original edition of Book I of the *Treatise* (London: John Noon, 1739), this paragraph is the first one beginning on 452. The footnote to the Appendix in the original edition of Book III (London: Thomas Longman, 1740), reads "Vol. I. page 452."

38. He expresses the same hopes at T 232 in I iv 5, a section that deserves more attention than I can give it here. There he says that although the intellectual world is "involv'd in infinite obscurities," it is "not perplex'd with any such contradictions, as those we have discover'd in the natural."

39. See, for example, Norman Kemp Smith, *The Philosophy of David Hume* (London: Macmillan, 1941), 558.

40. The two principles also play a positive role at T 207. We falsely suppose, Hume says there, that collections of related objects are simple and the same through time. This is at odds, he observes, with the principle that "every perception is distinguishable from another, and may be consider'd as separately existent." This principle is the conjunction of the two principles stated at T 636.

41. Although he defends an account of the Appendix very different from mine, Wade L. Robison makes an observation that captures my thought almost exactly. "Hume cannot explain why we think there is a self as a *trick* imposed on the mind by the nature of the relations between its impressions," he writes, "for there is nothing about the impressions that could constitute the basis for any trick." See "Hume on Personal Identity," *Journal of the History of Philosophy* 12 (1974): 190. Donald Ainslie emphasizes Hume's failure to account for the philosopher's commitment to the identity and simplicity of the self in "Hume's Reflections on the Identity and Simplicity of Mind." For passages emphasizing our commitment to simplicity and real identity see T 207, 255, and 635.

42. Here is a place where first-person pronouns are perhaps confined to philosophers, despite the absence of explicit indications. It is also a passage establishing a close link between I iv 6 and I iv 2. It adds to our puzzles: why, if Hume is unhappy with his explanation of our tendency to suppose real personal identity, is he not also unhappy with his explanation of our tendency to suppose continued existence?

43. Here Hume's concern with personal identity intrudes, but I do not think he is offering "substance" as just another label for what is picked out by "soul" or "self." He is pointing to a principle that accounts for *both* the fiction of material substance

and the fiction of immaterial substance. The latter fiction was the main theme of I iv 5, so Hume's mention of it here is not out of line.

44. The power of resemblance is also emphasized at T 204, and in the footnote spanning T 204–205, where Hume points out that there are *two* relations of resemblance at work in (i): the resemblance of the perceptions themselves, and the resemblance between the acts of mind that “survey” or reflect on those perceptions.

45. Hume is certainly saying here that (iii) and (iv) are dispensable. But he may also think that (ii) is dispensable, *in cases where we are confounding collections of perceptions, rather than single (and closely resembling) perceptions*. When it comes to a pair of closely resembling perceptions, it is always possible to find something invariable and uninterrupted—a single, unvarying perception—to justify the identity. The same can't be said when we are dealing with variable collections, as we are in the case of the mind or person.

46. I agree with Robison that by “the true idea of the human mind,” Hume means “the only idea we have, not that it is true that a person is the same from one moment to the next” (“In Defense of Hume's Appendix,” 93). But I take the causal relations at work in the passage to be *philosophical* rather than *natural*, which puts me at odds with the reading Robison defends on 94–98.

47. “Personal Identity and the Passions,” *Journal of the History of Philosophy* 27 (1989): 545–557. See in particular 551 and 553. McIntyre writes that “Book 2 of the *Treatise*, like Book I, takes causation to be the primary connection between the perceptions, thoughts, and actions that compose the self through time (553).”

48. At many points in both the *Natural History of Religion* and the *Dialogues concerning Natural Religion*, Hume (or one of his characters) speaks of nature as a single system or whole. In the *Treatise* itself, Hume attributes an extreme form of such a view to Spinoza, who is said to accept “the simplicity of the universe, and the unity of that substance, in which . . . both thought and matter . . . inhere” (T 240). In Spinoza's view, “neither time, nor place, nor all the diversity of nature, are able to produce any composition or change in its perfect simplicity and identity” (T 241).

49. At T 107–109, in explaining how we arrive at our belief in a “system of realities” (109) beyond those we sense and remember, Hume appeals to causation. But he uses it to account for our commitment to the *particulars* (see T 108) that populate the system. Our tendency to view them *as* a system is simply taken for granted.

50. At T 202–203, Hume describes resemblance as “of all relations . . . the most efficacious” in causing the mind to pass with facility from one idea to the other, but this is a bit misleading. It is “most efficacious” only because it plays a double role: as Hume explains (though again a bit misleadingly), “it not only causes an association of ideas, but also of dispositions, and makes us conceive the one idea by an act or operation of the mind, similar to that by which we conceive the other” (T 203). Causation plays only the first role, but it can then collaborate with resemblance—the resemblance between my state of mind when I consider an event and my state of mind when I consider its effect—to strengthen the original association. *And in the first role, causation can be more efficacious than resemblance*, as Hume in fact seems to acknowledge at T 11 (“there is no relation, which produces a stronger connexion in the fancy, and makes one idea more readily recall another, than the relation of cause and effect betwixt their objects”) and in early editions of *An Enquiry*

concerning *Human Understanding*, Section III, where he describes the relation of causation as "the strongest of all others." See Beauchamp, EHU 103.

51. Nidditch does take note of the excised passage; see David Hume, *Enquiries concerning Human Understanding and concerning the Principles of Morals*, ed. L. A. Selby-Bigge, 3rd ed. revised by P. H. Nidditch (Oxford: Clarendon Press, 1975), 358. The passage is now available in Beauchamp's edition, 102–107, and in Charles W. Hendel's much older edition, *An Inquiry concerning Human Understanding* (Indianapolis: Bobbs-Merrill, 1955), 33–39.

52. The unity of the work of art is a recurring theme in Shaftesbury's *Characteristicks*; the *Characteristicks* was, in fact, an especially influential statement of the neoclassical aesthetic ideal that Hume takes for granted in the excised portion of EHU Section III. See, for example, *Sensus Communis: An Essay on the Freedom of Wit and Humour*, I: 142–146 and C 66–67; MR 3: 24–25 and C 349; MR 3: 259–263 and C 448–449; MR 3: 263–264 and C 450; and "A Notion of the Historical Draught of Hercules," which was added to some copies of the *Characteristicks* in the 1714 edition. The "Notion of the Draught" appears in Shaftesbury's *Second Characters, or The Language of Forms*, ed. Benjamin Rand (Cambridge: Cambridge University Press, 1914), 29–61. There Shaftesbury commends the painting he is describing as "a single piece, comprehended in one view, and formed according to one single intelligence, meaning, or design; which constitutes a real whole, by a mutual and necessary relation of its parts, the same as of the members in a natural body" (32). The vocabulary is not unlike Hume's in the excised passage, and when Hume speaks of "that *Unity of Action*, about which all critics, after ARISTOTLE, have talked so much" (Beauchamp, EHU 103), Shaftesbury (who quotes extensively from the relevant passages in Aristotle's *Poetics*) is certainly among the critics he has in mind. Although I am not aware of any text in which Hume associates the unity of a beautiful object, natural or artificial, with the unity of a mind or person, this association is announced and reinforced throughout the *Characteristicks*. See, for example, the "Scale of BEAUTY" described in the long footnote at MR 3: 182–186; C 416–417.

53. The Appendix resembles I iv 6 in devoting more attention to identity than to simplicity. For example, Hume writes that all his hopes vanish when he comes to explain the principles "that unite our successive perceptions in our thought or consciousness" (T 636, emphasis mine). But at T 635 he does speak of his failure to explain what makes us attribute to our particular perceptions "a real simplicity" as well as a real identity.

54. Co-existent perceptions might, for example, have the same cause, or combine to yield a common effect. From the fact that I judge my parts are united-at-*t*, it does not follow that I make the judgment itself at *t*, or that when I do so, I consider only the direct relations that are visible at *t*.

55. Here it is worth observing that some perceptions last longer than others; see Jane L. McIntyre, "Is Hume's Self Consistent?" in Norton, Capaldi, and Robison, eds., *McGill Hume Studies*, 79–88. If a "moment" in my mental life is defined by the lifespan of a relatively stable perception (one lasting for several seconds), there is room even for direct causal relations among the more short-lived perceptions sharing (part of) that moment.

56. For illuminating discussion of this point see Ainslie, "Hume's Reflections on the Identity and Simplicity of Mind."

57. See William James's chapter on "The Self" in his *Psychology* (Greenwich, Conn.: Fawcett, 1963), 184, a passage I owe to Schechtman, who quotes it at *The Constitution of Selves*, 138: "There would be a hundred-and-first feeling there, if, when a group or series of such feelings were set up, a consciousness *belonging to the group as such* would emerge, and this one hundred and first feeling would be a totally new fact."

58. I am doubtful that Humean philosophers are committed to their own simplicity, but they may well feel a "propensity" (T 255) toward the commitment, and toward the conception of themselves as a "center" of their "different parts and qualities" (T 263). It can perhaps be argued that such a conception is at work in the passage I go on to quote from T 270–271, but in this paragraph I confine my attention to the Humean philosopher's commitment to his or her identity over time—time during which the principles inventoried at T 270–271 express themselves. I hope that in later work I will be able to describe more fully the Humean philosopher's commitment to, or propensity towards, the simplicity and identity of the self. What needs unpacking, I think, is Hume's conception of the self's "center" (T 263). Unlike, for example, a drama, which also unfolds over time, the self is somehow wholly present at every moment of its life. This presence may be the presence of its "center," which is supposed, even by the Humean philosopher, to remain strictly identical over time, despite a constant revolution in the "parts and qualities" surrounding it. *Treatise* I iv 5 may be helpful in understanding the challenges presented by such a supposition. Also relevant is a neglected observation at T 252: "When my perceptions are remov'd for any time, as by sound sleep; so long am I insensible of *myself*, and may truly be said not to exist." Here Hume embraces the Lockean view that sound sleep is perception-free (see also T 35). It follows (for Hume, though not for Locke) that the person who awakens in the morning is not the same as the person who falls asleep in the evening. I suspect that even the Humean philosopher thinks otherwise, at least on most days. This testifies to the tenacity of the fictions or tendencies that are at odds with Hume's metaphysics of the self.

59. A. H. Basson, *David Hume* (Harmondsworth: Penguin, 1958), 132–133, and S. C. Patten, "Hume's Bundles, Self-Consciousness, and Kant," *Hume Studies* 2 (1976): 59–64. For Garrett's assessment of this interpretation see *Cognition and Commitment*, 172–173.

60. Stroud, *Hume*, 134; see also 127.

61. For a wise and pithy summing-up of the interpretive challenge of the Appendix, see Robert J. Fogelin, *Hume's Skepticism in the Treatise of Human Nature* (London: Routledge, 1985), 108.