

**Raymond Martin**

Department of Philosophy  
Union College  
Schenectady, NY 12308  
martinr@union.edu

**Eighteenth Century British Theories of Self & Personal Identity**

1. In the *Essay*, Locke's most controversial claim, which he slipped into Book IV almost as an aside, was that matter might think (Locke1975:IV.iii.6;540-1).<sup>1</sup> Either because he was genuinely pious, which he was, or because he was clever, which he also was, he tied the denial that matter might think to the claim that God's powers are limited, thus, attempting to disarm his critics. It did not work. Stillingfleet and others were outraged. If matter can think, then for explanatory purposes the immaterial soul might be dispensable. But throughout the eighteenth century explanatory purposes were at the top of the agenda. And what had always made the soul so handy for proving immortality - that it is non-composite, static, and inaccessible to empirical examination - is also what made it so useless for investigating human nature. By contrast, consciousness is multifaceted, dynamic, and open to empirical investigation. Early in the century a Clarke might take the high road and resist descent into the merely probable and contingent. But when it came to investigating persons, the emerging science of human nature was the only game in town. One either played it or took oneself out of the action.

So far as personal identity itself is concerned, Locke had two main ideas, one negative and one positive. The negative idea was that the persistence of persons cannot be understood as parasitic upon the persistence of any underlying substance,

Chicago APA [21]

or substances, out of which humans or persons might be composed. The positive idea was that the persistence of persons can be understood only in terms of the unifying role of consciousness. Most of the time when Locke talked about consciousness in the context of talking about personal identity he meant *remembers*. His eighteenth century critics invariably attributed to him the view that a person at one time and one at another have the same consciousness, and hence are the same person, just in case the person at the later time *remembers*, from the inside, the person at the earlier time. But among the problems with this interpretation of Locke are, first, that there is reason to believe that Locke may not have been trying to present a non-circular analysis of personal identity over time, and second that one cannot on this interpretation explain what is central to Locke's view, that consciousness, which is reflexive, plays a dual role in self-constitution, unifying persons both *over* time and also *at* a time.

Locke's eighteenth century critics were right in thinking that the memory interpretation of personal identity that they attributed to him is vulnerable to decisive objections. However, almost all of them wanted to defeat the memory view in order to retain the view that personal identity depends on the persistence of an immaterial soul. Locke claimed that one can determine empirically whether someone retains the same consciousness over time, but not whether someone retains the same immaterial soul. As a consequence, he thought, the soul view is not only a wrong account of personal identity, but the wrong *kind* of account, whereas his own view, by contrast, is at least the right kind of account. For their part, Locke's early critics failed to see that even the memory view that they attributed to Locke was riding the crest of a wave of

naturalization that was about to engulf them. But as the century wore on their vision improved. Clarke's bravado, toward the beginning of the century, contrasts nicely with the subsequent defensiveness of Berkeley and Butler, a few decades later, and with the reluctance of most immaterial soul theorists, after Hume, even to do battle on the issue. And whereas toward the beginning of the century, it was enough simply to defend the immateriality of the soul and related a priori doctrines, such as the reflexivity of consciousness, without also contributing to the emerging science of human nature, eventually soul theorists tended to bracket their commitment to the soul in order to conduct meaningful empirical research. In the case of Berkeley, for instance, the immateriality of the soul is crucial to his metaphysics, but almost irrelevant to his inquiries into vision; and later, Hartley, Reid, and Tucker, who remained committed to the existence of the immaterial soul, similarly tended to segregate that commitment from their empirical inquiries.

As a consequence, in debates among theorists about the mind, as the century wore on it tended to matter less and less, what one's view was of the immaterial soul. Toward the end of the century, Hartley, the dualist, was an ally of Priestley, the materialist, while Reid, the dualist, attacked both. And while the main influences on Tucker, the dualist, were Locke, Clarke, and Hartley, it was not Locke and Hartley's dualism that most impressed Tucker, but their more scientific pursuits. It is only a slight exaggeration to suggest that Priestley could have put forth the very same views he did, even if, like Hartley, he had been a dualist; and Reid could have put forth most of his views, even if he had been a materialist.

2. Whereas Locke, in the *Essay*, sometimes used the words *person* and *self* interchangeably, he tended to use *self* to refer to a momentary entity and *person* to refer to a temporally extended one. Seemingly for other reasons, he defined the two terms differently (II.xxvii.9;335;17;341). His definition of *person* highlighted that persons are thinkers and, as such, have reason, reflection, intelligence, and whatever else may be required for trans-temporal self-reference. His definition of *self* highlighted that selves are sensors and as such feel pleasure and pain, and are capable of happiness, misery and self-concern. These differences in his definitions reflect disparate concerns that he expressed throughout his discussion of personal identity.

We know how, in Locke's view, humans come into being. It is a biological process. How do selves (or persons) come into being? His answer was that is a psychological process that involves five steps. It begins with a human organism's experience of pleasure and pain, which gives rise, first, to the idea of a self - its own self - that is the experiencer of pleasure and pain, and then to concern with the quality of that self's experience (each of us wants more pleasure, less pain). Then the momentary self thus constituted (or perhaps the organism) thinks of itself (or its self) as extended over brief periods of time (say, the specious present); finally, through memory and the appropriation ingredient in self-consciousness, it thinks of itself as extended over longer periods of time (Ibid.: II.xxvii.26; 346). Locke, thus, thought of the constitution of the self as at least being capable of being analyzed into an ordered, multi-step process. He may or may not have thought that the prior phases of this

process temporally precede the subsequent phases.

Whatever Locke's view on this question of timing, he clearly thought that self-constitution involves appropriation and that appropriation and accountability go hand in hand. A person, he said, is "justly accountable for any Action" just if it is appropriated to him by his self-consciousness (II. xxvii.16;341). He regarded the appropriation ingredient in self-consciousness as a natural relation between the organism and its present and past, which then is the basis for a non-natural relation of moral ownership. It is primarily this view - not, as some have suggested, a vague association between consciousness and conscience, or an equivocation on the notion of ownership - that connects the natural appropriation that is part of human psychology with the non-natural appropriation that is the concern of both ethics and the law.<sup>ii</sup>

**3.** One of the most puzzling aspects of Locke's account of personal identity is his view of the ontological status of persons. There are two aspects to the puzzle: his view of the status of *humans*; and his view of the status of *persons*. Commentators often assume that, in Locke's view, humans are substances, and the puzzle is that of determining whether persons are also substances. However, there is some reason to believe that in his chapter on identity Locke may have used the term substance in a more restricted sense than he did in the rest of the *Essay*. In this more restricted sense, only God, immaterial thinking things, and individual atoms would be particular substances. Cohesive collections of atoms - say, lumps of gold - would be collective substances. Other things that Locke elsewhere speaks of in a more expansive sense of

*substance* as particular substances - oak trees, horses and persons, for example - would, then, not be substances at all but, rather, particular mixed modes, that is, functional organizations of particular substances. If this interpretation is correct, then Locke was at least ambivalent about the substantial status of living things, including humans, and perhaps also of inanimate, macroscopic objects such as rocks and chairs.

Elsewhere in the *Essay*, Locke freely talks of living things as if they were substances. But it is unclear whether he means that they are substances in the same primary way as “Particles of Matter” are substances or are substances only in some second-class way and, hence, strictly speaking not really substances at all. Whatever Locke's intentions, he laid the groundwork for others to claim that selves are not substances, which then became a major issue. But even though Locke, perhaps unintentionally, encouraged the view that the self (or person) is a fiction, he clearly did not mean to suggest that it is fictional by being an artefact of legal or ethical theory. In his view, selves are created implicitly by human mentality via processes of appropriation and the application of self-concepts that are ingredient in reflexive consciousness.

**4.** In his chapter on identity, Locke was preoccupied with the implications of fission-like examples. He asked, for instance, "Could we suppose two distinct incommunicable consciousnesses acting the same Body, the one constantly by Day, the other by Night; and on the other side the same consciousness, acting by Intervals two distinct Bodies." Eventually he considered a case in which one's little finger is cut off and

consciousness, rather than staying with the main part of the body, goes with the little finger. He concluded that in such a case “’tis evident the little Finger would be the Person, the same Person; and self then would have nothing to do with the rest of the Body.” He later returned to this example, remarking that “though if the same Body should still live, and immediately from the separation of the little Finger have its own peculiar consciousness, whereof the little Finger knew nothing, it would not at all be concerned for it, as a part of it self, or could own any of its Actions, or have any of them imputed to him” (II.xxvii.18; 342). In this elaborated version of his example, Locke seems to be suggesting that the original consciousness splits, part of it going to the little finger and part remaining in the rest of the body, each part then constituting a whole consciousness qualitatively identical to the original. On this reading, his example is what philosophers in our own times would call a *fission example*, the first one to be considered explicitly in the context of personal identity theory. Since Locke did not explore the implications of his example it is impossible to know for sure whether this is what he had in mind. But once he published his new theory, the fission-example cat was out of the bag.

**5.** Between 1706 and 1709 Clarke and Collins confronted each other in a six-part written debate that by all indications was well-known throughout the century (Clarke 1738:v.3,720-913). Their point of departure was the question of whether souls are naturally immortal, where by “soul,” they agreed to mean “Substance with a Power of Thinking” or “Individual Consciousness” (750). Clarke defended the traditional Platonic

idea that souls are immaterial, and hence, indivisible and naturally immortal. Collins countered that the soul is material. Both agreed that individual atoms are not conscious. Their dispute turned on the question of whether it is possible that a system of matter can think. Clarke argued that it is not possible, Collins that matter does think. Throughout Clarke played the part of the traditional metaphysician, arguing largely on a priori grounds that the soul is indivisible, even though, in his view, it is extended. Collins, though not always consistently, played the role of the empirical psychologist. His faltering, but often successful attempts to reformulate traditional metaphysical issues empirically embodied the birth pangs of a new approach, one that grew steadily throughout the century. Their debate is, thus, a poignant record of two thinkers' struggles to cope with a rapidly changing intellectual climate, Clarke by hanging onto the old, Collins by groping for the new.

Collins' approach was the progressive side of Locke's, of whom he had been a close personal friend and disciple. Yet Collins was disposed and able to go beyond Locke, in two ways: first, methodologically, if not also ontologically, he was unabashedly materialist; second, he replaced Locke's metaphysically awkward same-consciousness view of personal identity with a more defensible connected-consciousness view. According to Collins, the basic problem with Clarke's account is that he was trying to settle by verbal fiat what should be settled empirically. Collins wanted instead "an account of what consciousness is," by which he meant a way of analyzing it empirically. In his view, different sorts of analyses would be required for several different aspects of consciousness (769-73).



In response to Collins' call for an empirical analysis of consciousness, Clarke countered mainly by reiterating a priori dogma. For instance, he claimed that strictly speaking, consciousness is neither a capacity for thinking nor actual thinking, "but the Reflex Act by which I know that I think, and that my Thoughts and Actions are my own and not Another's," and that "it would necessarily imply a plain and direct Contradiction, for any power which is really One and not Many . . . to inhere in or result from a divisible Substance" (784-7). In making such points, Clarke in effect invited Collins to explain how on an empirical view consciousness can be understood. However, in anticipation of Collins' response, Clarke conceded in advance that his own "affirming Consciousness to be an individual Power," was neither "giving an Account of what Consciousness" nor "intended to be so." It is enough, he continued, that "every Man feels and knows by Experience what Consciousness is, better than any Man can explain it" (790). It would in the end, then, come down to this: a clash between intuition and science.

As we know from similar debates in our own times, this particular conflict may be unresolvable. However, even in the eighteenth century it became clear that while intuition might be a sufficient basis to resist the reduction of the mental to the material, it was impotent as a source of explanations of mental phenomena. Collins returned to this point again and again (e.g., 807-9). He also claimed that in the case of remembering, he could explain how consciousness could be transferred from a material system of the brain initially composed of certain particles to one composed of other particles, without changing the individual subject of consciousness whose brain is

involved (809, 870). By our current standards, his explanation is crude. But it was a genuine scientific explanation, and Clarke had nothing comparable to offer.

What Clarke did instead was to accuse Collins of "begging the Question by, assuming the impossible Hypothesis" that the subject consisting of a brain which originally had an experience and a subject which subsequently remembered it might be the same. Clarke declared that on Collins' view, consciousness, rather than being a real individual quality, would be a "fleeting transferrable Mode or Power," and hence the self would be a fiction (844-5). Clarke assumed that pointing this out refuted Collins' view. Collins, for his part, appealed to an analogy between consciousness and the property of roundness to blunt the force of Clarke's claim that emergent properties cannot be "real." Collins pointed out that although individually the arches out of which a circle is composed are not round, collectively, when properly arranged, they are. Roundness, as even Clarke admitted, is a real property. In their final exchange, the question turned on whether the analogy between consciousness and roundness is a good one (860, 890, 894). Neither Clarke nor Collins had a principled, non-question-begging way of showing whether it was.

On the way to this inconclusive end, fission examples were introduced by Clarke, as a way of objecting to Collins' relational view of personal identity. Clarke introduced them by pointing out that if God in the afterlife can make one being with the same consciousness as someone who had lived on Earth, then God could make many such beings. Clarke took it as obvious that although in fact such multiple fission descendants would be different people, on Collins' view they would be the same person

(844-5, 852). Subsequently in their debate, fission examples were discussed several times. Because this debate was well-known, both fission examples and the idea that they have implications for personal identity theory were brought to the attention of eighteenth century theorists.

6. Joseph Butler, more than any other eighteenth century critic of Locke, took Locke's observations about the role of appropriation in self-constitution seriously. It is "easy to conceive," Butler said, "how matter, which is no part of ourselves, may be appropriated to us in the manner which our present bodies are" (1736: 86). But, he continued, where there is appropriation, there must be an appropriator. Locke had an appropriator in "man," which he distinguished from "person" and allowed might be merely a material organism. Butler thought that he (Butler) had already shown that the appropriator must be something simple and indivisible, and, hence, could not possibly be a material organism. This simple, indivisible appropriator, he assumed, is who we truly are. But what this being appropriates, he went on to explain, is not thereby part of itself, but, rather, something it owns. Butler had learned from Locke that, for all we know, the thinking principle in us may be material. So, he astutely conceded that the appropriator might be a simple material entity (87-8). In his view, it is our simplicity, not our immateriality, that ensures our survival. Butler thereby adapted the Platonic argument for immortality to the purposes of an age in which materialism was on the rise, recasting the a priori in an empirical mold.

When Butler turned to the topic of personal identity per se, he argued that on a

relational view such as that of Locke or Collins, people would have no reason to be concerned for the future life of the person who they nominally regarded as themselves (Ibid, 328). In response to what Butler saw as the dangers of empirical analysis, he proposed that we take as primitive the idea of personal identity, which he said, like the notion of equality, defies analysis. He added that we can determine intuitively that we have persisted, not just in “a loose and popular sense” such as we might employ in saying of a mature oak that it is the same tree as one that stood in its spot fifty years previously, even though it and that former tree have not one atom in common, but in “the strict and philosophical sense” which requires sameness of substance (330).

In Butler’s view, if our being were just to consist in successive acts of consciousness, then it would be a mistake “to charge our present selves with anything we did, or to imagine our present selves interested in anything which befell us yesterday; or . . . will befall us to-morrow; since our present self is not, in reality, the same with the self of yesterday, but another like self or person coming in its room, and mistaken for it: to which another self will succeed tomorrow” (331-2). On Locke's view, he claimed, we would have to consider ourselves to be selves and persons not really, but only in a fictitious sense. He thought that such a consequence refutes Locke's view. But he thought this not because he thought he could prove Locke's view is false (he admitted that he could not), but rather because "the bare unfolding this notion [that selves are merely fictitious entities] and laying it thus naked and open, seems the best confutation of it" (322, 325). Importantly because others, as we shall see, took a different view Butler’s death marked the end of an era in the main part of which religion

had dominated the philosophy of human nature.

7. When Locke published the *Essay*, he dreamt of the emergence of a science of human nature, but he was under the grip of several a priori assumptions that inhibited the development of that science. Chief among these was the Cartesian notion that all consciousness is reflexive. In addition, Locke had ulterior motives. Nothing is more central to his account of personal identity than his distinction between person and man (=human), but that distinction was an important one for him largely because it allowed him to suggest that matter might think while at the same time accommodating the Christian dogma of the resurrection. When Hume published *A Treatise of Human Nature* (1739), he talked as if a science of human nature had already emerged.

In developing his account of self and personal identity, Hume had gotten beyond both the a priori assumption that consciousness is reflexive and the dogma of the resurrection. Nevertheless, he too had a dream about the new empirical philosophy of human nature. His dream was not of its emerging but of its assuming its rightful position among the sciences, which, in his view, was at the foundation of a mighty edifice of human knowledge. Whereas today we tend to think of physics as the most fundamental science, Hume thought of the science of human nature as the most fundamental. His confusion on this issue was not a peripheral mistake, but central to his perspective. It was because he thought that the science of human nature - what today we would call psychology - includes philosophy that he thought it was not just another science, but the foundation of all the sciences. In his view, the science of human nature would not

only itself be founded on experience and observation, but - and this is how it would be the foundation of all the sciences - it would explain how all knowledge, including whatever is discovered in any of the other sciences, also is founded on experience and observation. As it happened, however, for psychologists to find their feet as scientists they had to abandon such epistemological and metaphysical pretensions and realize that it was not their job to get to the absolute bottom of things, a task that happily could be left to philosophers. As psychologists, it was their job to explain human behavior. To do that, they had to take certain things for granted that in a more philosophical frame of mind could be seen as deeply questionable. Fortunately, Hume did not stick consistently to his idea that the science of human nature would be the foundation of all the sciences. Sometimes, he seemed to see, if only through a glass darkly, that the new science would have a different mission.

The contrast between what I am calling, somewhat anachronistically, philosophical and scientific approaches is especially poignant in Hume's account of self and personal identity. In Book I of the *Treatise*, the heart of his account is his argument that belief in the substantial self is an illusion. More generally, he was intent on showing that belief in the persistence of anything is an illusion. This is what today we would call philosophy, rather than psychology. And in pursuing this philosophy, Hume, the skeptical metaphysician, is at his *destructive* best. However, in the remainder of Book I, Hume addressed the task of explaining why people are so susceptible to the illusion of self. And in Book II he explained how certain dynamic mentalistic systems in which we represent ourselves to ourselves, as well as to others, actually work, such as

those systems in us that generate sympathetic responses to others. In these more psychological projects, Hume took for granted many things that in Book I he had subjected to withering skeptical criticism. This is Hume the psychologist at his *constructive* best.

**[Beginning of Highlighted Material:**

In Hume's view, since all ideas come from impressions and there is no impression of a "simple and continu'd" self, there is no idea of such a self. This critique of traditional views led Hume to formulate his alternative "bundle" conception of the self and also to compare the mind to a kind of theatre, in which none of the actors - the "perceptions [that] successively make their appearance" - is the traditional self since none, strictly speaking, is either "simple" at a time or identical over time. Beyond that, Hume claimed, humans do not even have minds, except as fictional constructions. Thus, in Hume's view, a crucial respect in which minds are not analogous to real theatres is that there is no site for the mental performance, at least none of which we can have knowledge; rather, there "are the successive perceptions only, that constitute the mind; nor have we the most distant notion of the place, where these scenes are represented, or of the materials, of which it is compos'd" (Ibid, 253).

With these philosophical preliminaries out of the way, Hume turned to the psychological task of explaining how objects that are constantly changing, including the materials out of which we ourselves are constructed, nevertheless seem to persist. But before proceeding with this task he distinguished "betwixt personal identity, as it regards our thought or imagination, and as it regards our passions or the concern we

take in ourselves.” He added that the first of these issues - personal identity, as it regards our thought or imagination - “is our present subject” and that “to explain it perfectly we must take the matter pretty deep.” Among other things, that meant to him that we must account for the identity that we attribute to plants and animals, “there being a great analogy betwixt it, and the identity of a self or person” (Ibid).

The difference that Hume had in mind in distinguishing between his two ways of regarding personal identity is between, on the one hand, explaining why we regard anything that changes, including ourselves, as persisting over time (this is personal identity as it regards our thought or imagination) and, on the other, explaining the role that belief in ourselves as things that persist over time and through changes plays in the ways we represent ourselves to ourselves and to others (this is personal identity as it regards our passions or the concern we take in ourselves). The first of these occupies Hume in most of the remainder of Book I, the second in most of Book II.

In explaining personal identity as it regards our thought or imagination, the crucial psychological question for Hume was that of figuring out what causes us to forge a succession of perceptions into a persisting object. His answer, in one word, is: resemblance. When successive perceptions resemble each other, he said, it is easy to imagine that the first simply persists. In fact, “our propensity to this mistake” is so ubiquitous and strong “that we fall into it before we are aware.” And even when we become aware of our error “we cannot long sustain our philosophy, or take off this bias from the imagination.”<sup>iii</sup> Later Hume would claim - probably at least with Locke’s prince and cobbler example in mind - that “all the nice and subtile questions concerning



personal identity” are merely verbal. In the present context, however, he insisted that “the controversy concerning identity is not merely a dispute of words.” Usually, he continued, when people attribute identity “to variable or interrupted objects” their “mistake” is “attended with a fiction” (Ibid, 255). They believe that the identity, which they have claimed obtains, is not just their (perhaps pragmatically motivated) decision to regard distinct but similar objects as the same. Rather, they believe that those objects really are the same, perhaps even that what makes them the same is the existence of some unifying substance, such as soul, or some unifying mode, such as life or consciousness. Thus, in Hume’s view, normally it is not just that someone, in full knowledge of the facts, innocently chooses to call distinct objects which resemble each other the same object, but rather that the person who chooses to do this is immersed in a cloud of metaphysical confusion. Hume concluded this part of his discussion by comparing “the soul” to “a republic or commonwealth,” the seeming persistence of which is guaranteed by the relations among its parts, rather than by the persistence of any of its parts (Ibid, 261).

In Locke’s view, memory played a crucial role in constituting personal identity. In Hume’s view, it does so also, but for different reasons: It not only creates resemblances among successive perceptions, but also reveals to us that our perceptions are causally linked, information we then use as a basis for extending our identities to periods of our lives that we do not remember (261-2). In connection with the topic of forgetfulness, Hume said that, in his view, which presumably he intended to contrast with the views of Locke and perhaps also Collins, “memory does not so much produce as discover

personal identity, by shewing us the relation of cause and effect among our different perceptions” and that “it will be incumbent on those who affirm that memory produces entirely our personal identity, to give a reason why we can thus extend our identity beyond our memory” (262).

Hume extended his critique by questioning the seriousness of trying to make fine-grained distinctions, perhaps especially in the case of specially contrived, hypothetical examples, such as Locke’s prince and cobbler example, about whether personal identity obtains. He said, “Identity depends on the relations of ideas; and these relations produce identity, by means of that easy transition they occasion. But as the relations, and the easiness of the transition may diminish by insensible degrees, we have no just standard by which we can decide any dispute concerning the time when they acquire or lose a title to the name of identity.” It follows, he said, that “all the disputes concerning the identity of connected objects are merely verbal, except so far as the relation of parts gives rise to some fiction or imaginary principle of union” (Ibid.). In sum, Hume’s view seems to have been that disputes about identity are merely verbal, if they are about which relations, were they to obtain, would constitute identity. But the disputes are based on substantive mistakes, if the disputants suppose that what is merely successive is really the same. In any case, such disputes are always about fictitious imaginary constructs. In his view, that is all there is to say about identity over time and through changes.

Thus, Hume may have thought that a crucial difference between Locke and himself on the question of personal identity is that whereas Locke thought that there is

a fact of the matter about whether a person persists, Hume thought that there is a fact of the matter only about the circumstances under which the illusion of persistence is nourished. In his capacity as a psychologist, Hume tried to explain what those circumstances were. But he did not stop there. As soon as he moved on to the largely psychological concerns that dominate Book II of the Treatise, he became deeply involved in what today theorists would call social psychology of the self. In doing so, he abandoned, but probably without realizing that he had done so, the project of marching up directly to “the capital or centre of the sciences, to human nature itself.” In his capacity as social psychologist of the self, Hume returned to “the frontier,” thus completing a transition from skeptical philosophy to the most general sorts of associational issues, and then to specific psychological hypotheses about how self-representations function in our mental economy, as for instance in his explanation of how sympathy works.

In discussing personal identity, Hume never discussed fission directly, and he had little to say, and nothing new, about how personal identity might be analyzed in a way that links it to questions of accountability and interestedness. However, in his discussion of the example of a church that burns down and then is rebuilt, it seems that he may have been aware of the special problems for judgments of identity that arise in the case of fission. In claiming that, “without breach of the propriety of language,” we might regard the two churches as the same church even if the first was of brick and the second “of free-stone,” he added the caveat, “but we must observe, that in these cases the first object is in manner annihilated before the second comes into existence; by

which means, we are never presented in any one point of time with the idea of difference and multiplicity; and for that reason are less scrupulous in calling them the same” (Hume 1739: 258).

In Hume’s view, since all reason or understanding has to work with are diverse perceptions, it is the imagination that provides the links upon which our conception of self is ultimately based, even though people all but invariably create the fiction that they are something more than just perceptions imaginatively linked.<sup>iv</sup> Rather than considering the nature of personal identity per se, Hume turned instead, and almost exclusively, to two other questions: first, that of explaining how the fiction of identity arises, not only in the case of persons, but in that of anything which seems to persist over time and through changes; and, second, that of what role the fictional self plays in our emotions and motivations. He thus shifted the emphasis from conceptually analyzing the notion of personal identity to empirically accounting, first, for how it arises and, second, for its functional role.

**8.** Reid criticized Hume for supposing that there is nothing more to mind than a “succession of related ideas and impressions, of which we have an intimate memory and consciousness.” He asked “to be farther instructed, whether the impressions remember and are conscious of the ideas, or the ideas remember and are conscious of the impressions, or if both remember and are conscious of both? and whether the ideas remember those that come after them, as well as those that were before them?” His point was that since ideas and impressions are passive, they cannot do anything,

whereas Hume implied that the “succession of ideas and impressions not only remembers and is conscious” but also “judges, reasons, affirms, denies,” even “eats and drinks, and is sometimes merry and sometimes sad.” Reid concluded, “If these things can be ascribed to a succession of ideas and impressions in a consistency of common sense, I should be very glad to know what is nonsense.” Reid concluded that in any view in which substance has no place, agency would have no place either.<sup>v</sup> Since Reid thought it would be absurd to deny agency, substance had to be reintroduced. But whereas he assumed that the need for substance is an argument for immaterial substance, actually, so far as his argument goes, it shows at most only the need for substance of some sort. In any case, Reid, the immaterialist about the mind, here criticized Hume, the immaterialist about everything, for not being able to explain, on immaterialist grounds, the difference between impressions and ideas. This is one of the few places in Reid’s published work where his metaphysics of the soul may have made a substantive difference to the scientific account he was trying to develop.

In his own account of personal identity, Reid began by noting that “the conviction which every man has of his Identity, as far back as his memory reaches, needs no aid of philosophy to strengthen it; and no philosophy can weaken it, without first producing some degree of insanity.” Thus, “there can be no memory of what is past without the conviction that we existed at the time remembered.” He continued, “There may be good arguments to convince me that I existed before the earliest thing I can remember; but to suppose that my memory reaches a moment farther back than my belief and conviction of my existence, is a contradiction” (Ibid, 344). Although Reid did not specifically say

so, he seemed to suppose that, if we are rational, we automatically take ownership of the past thoughts, experiences, and actions that we remember. It seems, then, that Reid's continuing commitment to a reflexive account at least of memory, if not of all consciousness, may have prevented him from extending his new approach to a developmental account of the acquisition of self-concepts.

**9.** David Hartley was a methodological materialist, but not also a substantive one.

Differing in this respect from Collins before him and Priestley after, Hartley believed that "man consists of two parts, body and mind," where the mind "is that substance, agent, principle, &c. to which we refer the sensation, ideas, pleasures, pains, and voluntary motions." He accepted Locke's concession that it is possible, for all we know, that matter thinks. And he doubted that either problems with materialism or prescientific intuitions we may have about unity of consciousness could be used to prove that the soul is immaterial, confessing that "it is difficult to know [even] what is meant by the Unity of Consciousness." He did say, though, that the main problem with materialism is "that Matter and Motion, however subtly divided, or reasoned upon, yield nothing more than Matter and Motion still." But it was, he said, "foreign to [his] Purpose" to pursue the issue.

Hartley's humility about ontological questions extended to issues involving the afterlife. It is worth noting, he said, "that the Immateriality of the Soul has little or no Connexion with its Immortality; and that we ought to depend upon Him who first breathed into Man the Breath of the present Life, for our Resurrection to a better."<sup>vi</sup>

Hartley subsequently never used his metaphysical and theological views to determine the content of his more scientific views. He was a physico-theologian, in the tradition of Bacon, Boyle, Locke, and Newton before him, but in Hartley's scientific work it was the physico side that prevailed. In his case, like Locke's, his epistemological views engendered a deep humility, mixed with true religious piety, about the extent to which, through reason, humans can know metaphysical and religious truths.

However, in stark contrast to his metaphysical and theological humility, Hartley's associationist psychology was boldly speculative. There he claimed that all of human nature is built out of associations of sensations and consequent ideas, whose origins are in physical impressions in the organism. In his view, association was a principle in the service of the first truly general account of human and animal psychology, which, as it happened, laid the foundation for a mechanistic physiological psychology.

**10.** Joseph Priestley embraced materialism, with little concern about tradition, so long as he could maintain his own liberal version of Christianity. The latter was so liberal, that many Christians - including Reid - looked on him as a pariah, while atheists tended to accept him as one of their own. Yet, Priestley, like Hartley, was a paradigm of that fusion of theologian and scientist that flourished during the Enlightenment, particularly in England. He faced the facts of science and history head on, and modified his Christianity to make faith consistent with reason.

Like Hartley, Priestley was a gradualist. He saw the differences between humans and other animals as differences of degree, rather than kind. And, he saw human

infants as starting off more like other animals and only gradually learning adult human modes of thinking, including even the ability to conceptualize themselves. Priestley did not even mention the doctrine of the reflective nature of consciousness, which he merely assumed is wrong, an indication of how far the science of mind had progressed since Clarke's debate with Collins. And in Priestley's supposing that brutes and children differ from adult humans in not having second-order reflections, and that as a consequence their emotions are "less complex," one can glean how far the science of mind had progressed even since Hume (cp. Hume: 1739, 326). Finally, in Priestley's suggesting that children only gradually acquire self-concepts, he, in effect, invited others to explain how the notion of self is gradually acquired. His student, Hazlitt, accepted the invitation.

Priestley's rootedness in science, together with the matter of factness of his materialistic approach, differed radically from the epistemologically oriented way of ideas approach championed by Locke, Berkeley, and Hume. Although Priestley accepted "the way of ideas," he did not think that it led to skepticism about the external world, as Reid had claimed, or, indeed, to skepticism about anything. Priestley was a realist. He did not argue for the existence of an external world, beyond simply declaring that its existence is obviously the best explanation of the fact that different people report having similar experiences in the presence of a common stimulus. Since realism was so unproblematic for Priestley he made a much cleaner separation between philosophy and science than Hume, in particular, had been able to do. Priestley did not have a lot to say about personal identity and did not even discuss fission in his main



contribution to the personal identity debate, although he discussed it in correspondence with Price. Even so, his understanding of related issues that fed into Hazlitt's views and that have played an important part in the personal identity debates in our own times was more subtle and clear-headed than perhaps that of any other eighteenth century thinker.

Priestley's thoughts on personal identity are primarily his attempt, in response to criticisms from more conservative Christians, to show that his materialism is compatible with the Christian idea of resurrection. As follower of Hartley, he thought that the sentient and thinking principle in man must be "a property of the nervous system or rather of the brain" (Ibid). But he went further than Hartley in suggesting that the brain was not only necessary for human mentality, but sufficient as well. In Priestley's view, it is scientifically useless to postulate any immaterial substance to account for human behavior. Of course, all of this will sound quite modern to us. However, what is truly sophisticated and innovative in Priestley's treatment of personal identity is the way in his exploratory discussion of a hypothesis he downplays the importance of personal identity per se and highlights that of the functions that belief in our own identities actually serves.

Priestley began this part of his discussion by considering an objection, which he says was made to "the primitive Christians, as it may be at present" that "a proper resurrection is not only, in the highest degree, improbable, but even actually impossible since, after death, the body putrefies, and the parts that composed it are dispersed, and form other bodies, which have an equal claim to the same resurrection." He

continued: "And where, they say, can be the propriety of rewards and punishments, if the man that rises again be not identically the same with the man that acted and died?" In reply, he first makes it clear that in his opinion "we shall be identically the same beings after the resurrection that we are at present." Then, "for the sake of those who may entertain a different opinion," he proposes to "speculate a little upon their hypothesis" in order to show that "it is not inconsistent with a state of future rewards and punishments, and that it supplies motives sufficient for the regulation of our conduct here, with a view to it" (Ibid, 165). In other words, the task that Priestley sets himself is that of showing that even if after death "resurrected selves" [our term] are not strictly identical to anyone who existed on Earth it does not make any difference since identity is not what matters primarily in survival.

In arguing on behalf of this radical new idea, Priestley first endorsed Locke's view that so far as personal identity is requisite either for the propriety of rewards and punishments or for the concern that we take in our future selves, "the sameness and continuity of consciousness seems to be the only circumstance attended to by us." Then Priestley made it clear that, in his view, whether identity per se obtains is of no great consequence:

Admitting, therefore, that the man consists wholly of matter, as much as the river does of water, or the forest of trees, and that this matter should be wholly changed in the interval between death and the resurrection; yet, if, after this state, we shall all know one another again, and converse together as before, we shall be, to all intents and purposes, the same persons. Our personal identity will

be sufficiently preserved, and the expectation of it at present will have a proper influence on our conduct (Ibid, 166-7, emphasis added).

By Priestley's use of the expressions, "to all intents and purposes" and "sufficiently," he here separated the question of whether we will be identical with someone who exists in the future from that of whether it matters. In other words, what Priestley said here is that on the view under consideration, even if the resurrected person were not strictly identical with the person on Earth, he would be close enough to being identical that the loss of strict identity would not matter. And in considering whether strict identity does matter, he distinguished three ways in which it might: people's so-called self-interested concerns for their own futures; societal concerns that the prospect of future rewards and punishments motivate people to behave themselves; and theological concerns about the propriety of divine rewards and punishments. Thus, toward the end of the eighteenth century and perhaps without inferring anything from fission examples, Priestley introduced and sympathetically discussed one of the key ideas - that identity is not primarily what matters in survival - that has been central to the revolution in personal identity theory in our own times.<sup>vii</sup>

**End of Highlighted material.]**

**11.** William Hazlitt's first work, *An Essay on the Principles of Human Action*, was published in 1805, when he was twenty-seven years old.<sup>viii</sup> It was the culmination of a kind of perspective on personal identity that had begun with Locke and been developed by Collins, Hume, Law, and Priestley. Yet, with respect to certain questions that would become important in our own times, Hazlitt reads more like one of our own

contemporaries than any of his predecessors. He wrote that he was led to his central realizations by wondering "whether it could properly be said to be an act of virtue in anyone to sacrifice his own final happiness to that of any other person or number of persons, if it were possible for the one ever to be made the price of the other?"

Suppose that one could save twenty other persons by voluntarily consenting to suffer for them. "Why," he asked, "should I not do a generous thing, and never trouble myself about what might be the consequence to myself the Lord knows when?"

On behalf of common sense, Hazlitt answered that "however insensible" he may be now to his own interest in the future, when the time comes he shall feel differently about it and "shall bitterly regret my own folly and insensibility." So, he continued, still replying on behalf of common sense, "I ought, as a rational agent, to be determined now by what I shall then wish I had done, when I shall feel the consequences of my actions most deeply and sensibly. It is this continued consciousness of my own feelings which gives me an immediate interest in whatever relates to my future welfare, and makes me at all times accountable to myself for my own conduct" (Ibid, 133-35). Hazlitt was dissatisfied with this common sense answer.

I cannot . . . have a principle of active self-interest arising out of the immediate connection between my present and future self, for no such connection exists or is possible. I am what I am in spite of the future. My feelings, actions, and interests must be determined by causes already existing and acting, and are absolutely independent of the future (135).

Where there is no "intercommunity of feelings," he claimed, "there can be no identity of

interests" (139).

Hazlitt conceded that because we remember only our own past experiences and are directly "conscious" only of our own present experiences, in relation to the past and present people are naturally self-interested (110-1). The reasons for this, he said, are physiological. Memories depend on physical traces of prior sensations, and these traces are not communicated among individuals. Present sensations depend on the stimulation of one's nerves, and "there is no communication between my nerves, and another's brain, by means of which he can be affected with my sensations as I am myself." In the case of the future, however, Hazlitt stressed that people are neither "mechanically" nor "exclusively" connected to themselves. They cannot be, he thought, since no one's future yet exists. Instead, people are connected both to their own futures and to the futures of others by anticipation, which unlike memory and sensation, is a function of imagination and, thus, does not respect the difference between self and other.<sup>ix</sup> He maintained that to feel future-oriented concern for someone, one first must project oneself imaginatively into the feelings of that person, and imagination, functioning "naturally," that is, independently of its having acquired a bias through learning, projects as easily into the feelings of others as into one's own future feelings.

Hazlitt no doubt exaggerated the extent to which memory is independent of imagination and underestimated our mechanical connections to our future selves. For instance, in claiming that "I am what I am in spite of the future," he ignored the possibility that future person-stages of himself are none the less person-stages of himself for being future. Even so, he was right in insisting that at any given time there is

a crucial difference in our relations to our past and future selves. It is that we are already affected by past stages of ourselves and not yet affected by future stages. Hence, at any given time, our imaginations play a greater role in linking current to future stages of ourselves than to past stages of ourselves. And he may also have been right in insisting that so far as our values are concerned, past, present, and future do not have the same status or, if they do, that they come to have it in different ways. He claimed that to understand what these different ways are, one must investigate two issues that none of his predecessors had addressed adequately. One is the role of the imagination in connecting us to the future generally and, in particular, to ourselves in the future; the other is the role of self-conceptions in possibly masking from ourselves salient differences between past, present, and future.

It was in Hazlitt's account of the role of self-conceptions in our values and in our views of our own interests that he contrasts most sharply with the eighteenth century tradition of which he was the culmination. According to him, people are naturally concerned about whether someone is pleased or suffers as a consequence of their actions. This is because "there is something in the very idea of good, or evil, which naturally excites desire or aversion." But, he wrote, before the acquisition of self-concepts, people are indifferent about whether those who may be pleased or suffer are themselves or others: "a child first distinctly wills or pursues his own good," he said, "not because it is his but because it is good." As a consequence, he claimed, "what is personal or selfish in our affections" is due to "time and habit," the rest to "the principle of a disinterested love of good as such, or for its own sake, without any regard to

personal distinctions" (33-4). He thought that such thoughts provided a basis for founding morality not on self-interest, which he regarded as an "artificial" value, but on the natural concern people have to seek happiness and avoid unhappiness, regardless of whose it is (48-9).

Hazlitt's adopting this perspective prompted him to ask a question which did not arise as starkly or in the same form for any of his predecessors. The question was: If people connect to the future through imagination, which does not respect the difference between self and other, why is the force of habit almost invariably on the side of selfish feelings? His answer involved his trying to account for the growth of selfish motives in humans by appeal to their acquisition of self-concepts. In his view, when very young children behave selfishly it is not because they like themselves better, but because they know their own wants and pleasures better. In older children and adults, he thought, it is because they have come under the control of their self-concepts, which is something that happens in three stages. First, young children acquire an idea of themselves as beings who are capable of experiencing pleasure and pain. Second, and almost "mechanically" (since physiology insures that children remember only their own pasts), children include their own pasts in their notions of themselves. Finally, imaginatively, they include their own futures (34-5). The first two of these stages may have been suggested to Hazlitt by his reading of Locke. Arguably the third is original. However, even in the case of the first two, Hazlitt thought of them less as a philosopher and more as a psychologist might think of them, in terms of the acquisition of self-concepts, and whereas it was unclear whether Locke meant to distinguish

developmental stages in the acquisition of self-concepts, Hazlitt clearly meant to.

In the first half of the eighteenth century, the possibility of a developmental account of the acquisition of self-concepts that Locke may have seen dimly were invisible to most of his readers. As commonsensical as the idea of this sort of psychological development may seem to us today, it did not begin to emerge in the views of eighteenth century thinkers until mid-century. Hartley had a developmental, associational account of the mind, but he focused on the development of the passions and did not consider the acquisition of self-concepts. Rousseau, in *Emile*, was sensitive to developmental concerns, but not particularly with respect to the acquisition of self-concepts. Reid, late in the century, had a developmental psychology, but because of his commitment to the immateriality of the soul and the reflexive nature of consciousness, he may actually have made an exception in the case of the idea of self. Priestley, largely under the influence of Hartley, did think that his own developmental account could be extended to the acquisition of self-concepts, but he did not elaborate.

Hazlitt thought that to progress through all three of the development stages that he distinguished in the acquisition of self-concepts, a child has to differentiate its own mental activities from those of others. In his view, this involves "perceiving that you are and what you are from the immediate reflection of the mind on its own operations, sensations or ideas." He then raised the question of how a child's formation of self-concepts is related to its development of empathy and sympathy. No one previously had ever asked this question.<sup>x</sup>

In Hume's emotional contagion model of human sympathy, humans infer from



external behavior, facial expressions, and the like that others are in some particular mental state. Then, the resulting idea that humans form of another's state becomes converted in their own minds into an impression, so that now they too are in the same state, though perhaps less vivaciously. In explaining how this conversion from idea to impression occurs, Hume appealed to the idea's "proximity" in one's mind to the impression one has of oneself, which he said is "so lively" that "it is not possible to imagine that any thing can in this particular go beyond it." But, then, he added not a word of explanation about how people acquire their super-lively self-impressions.

Two decades later, Adam Smith gave an unusually thorough account of the role, in sympathy, of shifts from one's own to another's point of view. Yet Smith never attempted to explain how people acquire their ideas of the distinction between self and other. Aside from the applications of his ideas to ethical theory, his gaze was fixed on the importance of point of view as a feature of adult minds, not on the psycho-genetics of point of view in our mental development. The closest Smith came to discussing the mentality of children was in his explanations of how adults sympathize with "poor wretches," children, and the dead. In his view, in so sympathizing, adults do not simply replicate the other's state of mind in their own minds but, rather, imagine what they themselves would feel if they were reduced to the other's situation, but somehow, per impossible, allowed to keep their own current reason and judgment. In short, whereas Smith was preoccupied with explaining how sympathy is possible, it did not occur to him to explain how the conceptual apparatus that makes it possible came to be acquired in the first place.

Hazlitt speculated that young children imaginatively include only their own futures and not the futures of others in their ideas of self because the "greater liveliness and force" with which they can enter into their future feelings "in a manner identifies them" with those feelings. He added that once the notion of one's own personal identity is formed, "the mind makes use of it to strengthen its habitual propensity, by giving to personal motives a reality and absolute truth which they can never have" (140). This happens, he thought, because "we have an indistinct idea of extended consciousness and a community of feelings as essential to the same thinking being," as a consequence of which we assume that whatever "interests [us] at one time must interest [us] or be capable of interesting [us] at other times" (10-1).

Hazlitt claimed that a bias in favor of ourselves in the future could never "have gained the assent of thinking men" but for "the force" with which a future-oriented idea of self "habitually clings to the mind of every man, binding it as with a spell, deadening its discriminating powers, and spreading the confused associations which belong only to past and present impressions over the whole of our imaginary existence." However, whereas a host of previous thinkers - Descartes, Locke, Berkeley, Butler, and others - thought that people have intuitive knowledge of their own identities, Hazlitt rejected as "wild and absurd" the idea that we have an "absolute, metaphysical identity" with ourselves in the future, and hence that people have identities that are available to be intuited. We have been misled, he claimed, by language: by "a mere play of words." In his view, both children and adults fail to look beyond the common idioms of personal identity and as a consequence routinely mistake linguistic fictions for metaphysical

realities. To say that someone has a "general interest" in whatever concerns her own future welfare "is no more," he insisted, "than affirming that [she] shall have an interest in that welfare, or that [she is] nominally and in certain other respects the same being who will hereafter have a real interest in it." No amount of mere telling "me that I have the same interest in my future sensations as if they were present, because I am the same individual," he claimed, can bridge the gulf between the "real" mechanical connections I have to myself in the past and present and the merely verbal and imaginary connections that I have to myself in the future (6, 10-1, 27-9).

Assuming that people have no mechanical connections to themselves in the future, it follows, Hazlitt thought, that so far as peoples' "real" interests are concerned, their "selves" in the future are essentially others. So, for instance, if you've injured yourself, you may in the present suffer as a consequence. But "the injury that I may do to my future interest will not certainly by any kind of reaction return to punish me for my neglect of my own happiness." Rather, he concluded, "I am always free from the consequences of my actions. The interests of the being who acts, and of the being who suffers are never one." So, it makes no difference "whether [you] pursue [your] own welfare or entirely neglect it" (31). Your suffering in the future is only nominally your suffering.

In sum, Hazlitt gave a psychological account of how people come to identify with their future selves, from which he drew a metaphysical conclusion: that peoples' seeming identities with their future selves are based on an illusion. He then used this metaphysical conclusion as the basis for an inference to a normative conclusion: that

we have no self-interested reason to be concerned about the fate of our future selves. Whether or not Hazlitt's conclusions are correct, both of these inferences of his are fallacious. Psychology, all by itself, does not have such implications for metaphysics, and metaphysics, all by itself, does not have such implications for ethics. Nevertheless, Hazlitt asked questions, perhaps for the first time, that have deeply interested some theorists in our own times.

Hazlitt's consideration of fission examples occurred in the context of his critique of the Lockean idea that one's identity extends as far as one's consciousness extends. What, Hazlitt asked, would a theorist committed to this idea say "if that consciousness should be transferred to some other being?" How would such a person know that he or she had not been "imposed upon by a false claim of identity?" He answered, on behalf of the Lockeans, that the idea of one's consciousness extending to someone else "is ridiculous": a person has "no other self than that which arises from this very consciousness." But, he countered, after our deaths:

this self may be multiplied in as many different beings as the Deity may think proper to endue with the same consciousness; which if it can be so renewed at will in any one instance, may clearly be so in a hundred others. Am I to regard all these as equally myself? Am I equally interested in the fate of all? Or if I must fix upon some one of them in particular as my representative and other self, how am I to be determined in my choice? Here, then, I saw an end put to my speculations about absolute self- interest and personal identity (135-6).

Thus, Hazlitt saw that, hypothetically, psychological continuity might not continue in a

single stream but instead might divide. In asking the two questions - "Am I to regard all of these [fission descendants] as equally myself? Am I equally interested in the fate of all [of these fission descendants]?" - he correctly separated the question of whether identity tracks psychological continuity from that of whether self-concern tracks it. And, in direct anticipation of what would not occur again to other philosophers until the 1960s, he concluded that because of the possibility of fission neither identity nor self-concern necessarily tracks psychological continuity.

Hazlitt also used fission examples to call into question whether in cases in which there is no fission, a person's present self-interest extends to his or her self in the future. He began by asking:

How then can this pretended unity of consciousness which is only reflected from the past, which makes me so little acquainted with the future that cannot even tell for a moment how long it will be continued, whether it will be entirely interrupted by or renewed in me after death, and which might be multiplied in I don't know how many different beings and prolonged by complicated sufferings without my being any the wiser for it, how I say can a principle of this sort identify my present with my future interests, and make me as much a participator in what does not at all affect me as if it were actually impressed on my senses?

Hazlitt's answer was that it cannot.

It is plain, as this conscious being may be decomposed, entirely destroyed, renewed again, or multiplied in a great number of beings, and as, whichever of these takes place, it cannot produce the least alteration in my present being -

that what I am does not depend on what I am to be, and that there is no communication between my future interests and the motives by which my present conduct must be governed.

He concluded:

I cannot, therefore, have a principle of active self-interest arising out of the immediate connection between my present and future self, for no such connection exists, or is possible. . . . My personal interest in any thing must refer either to the interest excited by the actual impression of the object which cannot be felt before it exists, and can last no longer than while the impression lasts, or it may refer to the particular manner in which I am mechanically affected by the idea of my own impressions in the absence of the object. I can therefore have no proper personal interest in my future impressions. . . . The only reason for my preferring my future interest to that of others, must arise from my anticipating it with greater warmth of present imagination (138-40).

With the exception of F.H. Bradley, such ideas would not be taken seriously again until the late 1960s. Hazlitt not only conceded but embraced and celebrated the idea that the self is a fictional construct, since, in his view, this idea had the further implication that people have no special (“self-interested”) reason to value their future selves. At least to his own satisfaction, and in a way that clearly anticipated the work of Derek Parfit and others in our own times, Hazlitt tried to explain how the idea that the self is a fiction, far from being destructive to theories of rationality and ethics, actually made them better. In the process, he sowed the seeds, albeit on barren ground, of a modern

psychology of the acquisition of self-concepts and of a modern approach to separating the traditional philosophical problem of personal identity from the question of what matters in survival.

Hazlitt was the last progressive figure in a more or less continuous tradition of discussion of the nature of self and personal identity that began with Locke and that took place in Britain throughout the eighteenth century. Two things were mainly responsible for interrupting this tradition of discussion: One of these was the newly emerging separation of philosophy and psychology, each of which throughout the nineteenth and increasingly into the twentieth centuries tended to go their separate ways. Another was Kant, whose views only began to be felt seriously in Britain at the beginning of the nineteenth century, but once felt effectively changed the focus of debate about the self.\*

#### ENDNOTES

\*The present paper is based on research done with John Barresi, the results of which are published in *Naturalization of the Soul: Self and Personal Identity in the Eighteenth Century* (London: Routledge, 1999). Most of my claims in the present paper are elaborated and defended at greater length in that book. I am grateful to John for allowing me to borrow from this work that we did together

---

i. (Locke1975:IV.iii.6;540-1) = John Locke, An Essay Concerning Human Understanding, Bk. IV, Ch.3, section 6, which in the paperback version of Peter H.

---

Nidditch's edition of the Essay is on pp. 540-1 (Oxford University Press; hardback, 1975; paperback, 1979). Subsequent parenthetical references to Locke's Essay should be understood similarly.

ii. See Michael Ayers, *Locke*, 2 vols. (London: Routledge, 1991), v. 2, pp. 266-67. Why did Locke think that appropriation and accountability go hand in hand? There is an admittedly speculative but still, I think, plausible reason why Locke might have thought this. He might have thought that, analogous to the way in which people come under the rule of their government by constituting themselves as a body-politic, and hence accepting civil responsibility for what they do, humans become subject to ethical norms by constituting themselves as persons and hence accepting ethical responsibility for what they do. That is to say, humans, merely in virtue of being alive, and, hence, in virtue of being humans, do not, as it were, accept accountability for their pasts. But humans (or, persons) do accept accountability for their pasts, or at least for those parts of their pasts that they remember, when, through consciousness, they declare ownership of the various parts that collectively constitute themselves.

iii. In and of itself, Hume suggested, our supposing that objects persist is not so bad. But "in order to justify to ourselves this absurdity," we make up a story, often one in which the principle character is the notion of substance; that is, we invent the fictions of "soul, and self, and substance to disguise the variation" in our perceptions. When, as in the case of "plants and vegetables," we cannot fool ourselves into believing that the



---

persistence of an underlying substance accounts for the persistence of the organism, we invent an equally “unknown and mysterious” surrogate - presumably, “life” - to connect the successive and different perceptions (*Treatise of Human Nature*, 1739, L.A. Selby-Bigge, ed., Oxford: Clarendon Press, 1888, pp. 254-55).

iv. "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 and uninterrupted, or of something mysterious and inexplicable, or at least with a propensity to such fictions" (Ibid.: 255). And further: "The identity which we ascribe to the mind of man is only a fictitious one, and of a like kind with that which we ascribe to vegetable and animal bodies. It cannot, therefore, have a different origin, but must proceed from a like operation of the imagination upon like objects" (Ibid.: 259).

v. *Essay on the Intellectual Powers of Man*. In W. Hamilton, ed., *Philosophical Works of Thomas Reid* Vol. 1 (213-508), reprinted Hildesheim: George Olms, 444.

vi. *Observations on Man, His Frame, His Duty, and His Expectations*, 1749, two vols., reprinted 1966 (two vols in one), with an introduction by T. L. Huguelet, Gainesville, FL: Scholars' Facsimiles & Reprints, 512.

vii. In *Naturalization of the Soul*, Barresi and I inadvertently may have suggested that Priestley committed himself to the view here under discussion. He did not.

---

viii. *An Essay on the Principles of Human Action and some Remarks on the Systems of Hartley and Helvetius*, 1805, reprinted, 1969, with an introduction by J.R. Nabholtz, Gainesville, FL: Scholars' Facsimiles & Reprints.

ix. “[Imagination] must carry me out of myself into the feeling of others by one and the same process by which I am thrown forward as it were into my future being and interested in it. I could not love myself, if I were not capable of loving others. Self-love, used in this sense, is in its fundamental principle the same with disinterested benevolence,” Ibid, 3.

x. Currently developmental psychologists are preoccupied with this question. For a review of the literature, see Barresi, J. and Moore, C. (1996) ‘Intentional relations and social understanding’, *Behavioral and Brain Sciences* 19:107-54. In this paper, the authors argue that learning about self and other are co-occurring complementary functions, of which involve imagination or empathy. On the way to this conclusion they review relevant developmental literature on the theory of mind, including that on the early stages of sympathy.