

# 10

## Consciousness and personal identity

*Donald Ainslie and Owen Ware*

The “new science” of the seventeenth century was, broadly speaking, mechanistic, in that it conceived of the operations of nature as akin to the operations of artificial machines, the parts of which produce their effects through contact motion. Philosophers wrestled with the question of how human beings fit into this new understanding of nature. Some – notably René Descartes – insisted that the human soul was an immaterial, immortal, thinking substance that was entirely different from the extended world of the natural philosophers. Others – notably Thomas Hobbes – took human beings to be fully composed from deterministic matter in motion. Eighteenth-century philosophers struggled to reconcile these two poles of seventeenth-century thought, especially in light of the acknowledged explanatory limitations of the Newtonian version of the new science. Their starting point was John Locke’s revolutionary discussion of persons in the second edition of his *Essay concerning Human Understanding* (1694 ed., Locke 1689). By holding that human beings have only limited insight into the deep structures of both matter and mind, at the same time as they have a full grasp of themselves, Locke introduces the concept of persons as subjects of consciousness who also simultaneously recognize themselves as such subjects. Whether persons ultimately reside in immaterial substances or are simply elements of material systems can be left as an open question. Locke thought that his account of persons sufficed for our everyday concern with ourselves and our practices of moral attribution.

The problem is that Locke is less than clear about how his readers should understand his crucial innovation – his notion of consciousness. The difficulties were apparent from the start. Pierre Coste, the French translator of the *Essay*, in an edition originally overseen by Locke himself, notes that he cannot find an appropriate French word for “consciousness,” finally settling on “*con-science*” despite registering reservations about its appropriateness (Locke 1742: 1005). And G. W. Leibniz, in his response to Locke, *New Essays concerning Human Understanding*, introduces a French neologism, “*consciosité*,” out of a sense that “*con-science*” failed to capture the distinctive sense Locke intended (Leibniz 1765: 235n). Samuel Clarke lists five different possible meanings for “consciousness”: the reflexivity whereby all thoughts declare

DONALD AINSLIE AND OWEN WARE

themselves to belong to their subject (“the strict and properest Sense of the Word”), the act of thinking, the power to think, simple sensation, or the capacity to determine motion by the will (Clarke 1706: 177n). It is thus essential to read eighteenth-century philosophers with care, not assuming that an invocation of consciousness is self-explanatory.

We start the discussion below with an overview of Locke’s view that persons are self-conscious subjects. Locke’s initial respondents tended to reject his new notion of persons, insisting that a traditional understanding of persons as immaterial souls or substances was required to secure moral responsibility, both in this life and after. In his *Treatise of Human Nature* (1739–40), David Hume breaks with both Locke and the traditionalists by rejecting both substance ontology and Locke’s conception of consciousness. Whereas Locke emphasizes the *inward* character of mental life, with self-awareness an element of all thoughts, Hume takes us normally to lack any such sense of self. And thus, for him, an account of the mind was inadequate as an account of our everyday self-concern. Minds are nothing but bundles of perceptions, lacking intrinsic unity at a time or across time; nonetheless, our emotional responses to one another mean that persons in everyday life are defined by their virtues, vices, bodily qualities, property, riches, and the like.

Jean-Jacques Rousseau also takes persons to be fundamentally determined by our socially mediated emotional responses to one another, though unlike Hume or Locke, he has little interest in placing this account of persons alongside a larger discussion of the human mind and its operations. Rousseau’s focus instead is on the psychic and social instability that our emotional self-definitions create. We ultimately need to reform society itself to allow us to live in peace with ourselves and others, no longer threatened by our tendency to be defined by others’ reactions to us.

We finish the discussion with a consideration of Immanuel Kant’s complex and multifaceted treatment of the person. Like Hume, he thinks that we do not *experience* ourselves as conscious subjects in the manner that Locke suggested. But Kant’s innovation is that we nonetheless *presuppose* a special kind of self-consciousness in all of our judgments. We *can* think of ourselves as free in so judging and our moral commitments require that we *must* take ourselves to be free. The fundamental equality that Rousseau sought in the political order is, for Kant, a requirement that reason puts on all of us.

### Locke and the conscious person

Locke’s empiricism requires that every simple idea arise either from sensation or from introspective observation of our mental operations, what he calls “reflection” or “internal sense” (Locke 1689: 2.1.2–5, 104–6). His distinctive account of persons arises because neither source yields a full understanding of substances, mental or physical. We sense an object’s shape, color, heaviness, feel, and so on, but we do not similarly sense that in virtue of which these different qualities are united into one thing (*ibid.*: 2.23.4, 297). Similarly, we reflectively observe the mind’s perceiving ideas, compounding them into complex ideas, willing bodily or mental actions, and so on, but do not observe the mental substance in virtue of which these different

## CONSCIOUSNESS AND PERSONAL IDENTITY

mental features are united into one thing (ibid.: 2.23.5, 297–98). Locke thinks that our ignorance of matter and mind is so great that we are even unsure whether they are ultimately different from one another. Notoriously, he allows that it is possible that God “superadded” thought to matter so that human beings would be merely material systems (ibid.: 4.3.6, 539–43). Locke’s exact point has been debated extensively in the secondary literature (Ayers 1981; Downing 1998; McCann 1994; Rozemond and Yaffe 2004; Wilson 1979), and we need not dwell on it here. But his skepticism about our insight into substances, both mental and material, sets the stage for his discussion of persons. For, though we lack a clear grasp of material or immaterial substances, we are nonetheless able to recognize our own identity through time. Thus we must take persons to be something other than straightforward substances.

Locke discusses persons in a chapter, “Of identity and diversity” (Locke 1689: 2.27, 328–48), that he added to the *Essay* for its second edition in 1694. He argues there that an inanimate, natural, material object is the same through time so long as it neither gains nor loses any of its constituent particles. Artifacts, in contrast, remain identical through changes in their constitution so long as they preserve their function. Living beings, similarly, can sustain changes – sometimes radical changes, as when an acorn grows into an oak – while remaining the same so long as they preserve their organization or “life” (ibid.: 2.27.5, 331–32). Humans, for example, remain the same even if they lose consciousness or undergo radical changes in personality, so long as they are the same living beings (ibid.: 2.27.6, 330–31). But Locke thinks that we can imagine, say, a prince and a cobbler switching bodies, so that the cobbler’s memories and mental outlook would be located in the prince’s body and vice versa. No one would say that the prince *as a human being* is different because he is now possessed, as it were, by the cobbler. But neither would we say that, *as a person*, the prince remains in his original body; instead we would address the cobbler to find out what the prince thinks and desires (ibid.: 2.27.15, 340). Persons are thus different from human animals.

Locke similarly distinguishes persons from spiritual substances or souls. For he points out that, for all we know given our fundamental ignorance of substances, someone might have the same *soul* as the one Socrates once had. But, lacking any memories of life in the *agora*, of teaching Plato, and so on, we would not thereby say that he was the same *person* as Socrates (1700 ed., Locke 1689: 2.27.14, 338–39). Conversely, given our ignorance, it is possible that several substances have been supporting someone’s mental life, with each substance passing on to the next all of its memories and consciousness. In that case, the person would be the same while her or his soul varied (ibid.: 2.27.13, 337–38).

Thus Locke concludes that a person can be equated neither with a human animal nor with a soul. Instead a person is

a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing, in different times and places; which it does only by that consciousness, which is inseparable from thinking, and, as it seems to me essential to it: It being impossible for any one to perceive, without perceiving, that he does perceive. When we see, hear, smell, taste,

DONALD AINSLIE AND OWEN WARE

feel, meditate, or will any thing, we know that we do so. ... And as far as this consciousness can be extended backwards to any past Action or Thought, so far reaches the identity of that *Person*; it is the same *self* now it was then; and 'tis by the same *self* with this present one that now reflects on it, that that Action was done.

(Locke 1689: 2.27.9, 335)

A person is a self-conscious rational subject of thought and action. Moreover, Locke thinks that this self-consciousness brings self-concern in its wake: “a concern for Happiness [is] the unavoidable concomitant of consciousness; that which is conscious of Pleasure and Pain, desiring that that *self*, that is conscious, should be happy.” “Person” is thus a “Forensick” term (ibid.: 2.27.26, 346). In rewarding or punishing people for good or bad deeds we look not to the presence of an ongoing substance or human being, but rather to a connection between the consciousness of the one who did the deed and the one who is being held accountable. In human “Judicatures” we are often forced to rely on the identity of the human being as a proxy for the human person, but God in the afterlife ensures that we are properly rewarded or punished only for those deeds we did as conscious persons (ibid.: 2.27.22, 343–44).

In defining “person” in these terms, Locke is often thought to embrace a memory criterion for personal identity, whereby someone is the same as a person at an earlier time if and only if the later person remembers the earlier person’s experiences. And indeed Locke does repeatedly appeal to a person’s memory of her or his past throughout his discussion. But a memory criterion is open to serious objections both because it seems possible for someone to have forgotten (even irremediably so) previous experiences without thereby no longer having done them, and because it seems possible for someone to have false memories without thereby becoming the (non-existent?) person who did the deed that is supposedly remembered. Locke’s critics argue that we need to return to the traditional notion of persons as immaterial substances.

Thomas Reid, for example, worries that Locke’s view would violate the transitivity of identity. A brave officer might remember the experiences he had as a young boy and a general “in advanced life” might remember the experiences of the brave officer, but the general might have forgotten his experiences as a boy. If Locke held a memory criterion, then the officer would be the same person as the boy and the general would be the same person as the officer, but the general would not be the same as the boy (Reid 1785: 3.6; 276).

Both Leibniz and Joseph Butler also take Locke to hold a memory criterion for personal identity and both suggest that he has confused the evidence we appeal to in determining personal identity with its metaphysical foundation. Leibniz takes rational souls to include consciousness or “apperception” – a notion that later German philosophers, notably Kant, would develop in greater detail – of some of our perceptions, though he also allows that we have many “*petites perceptions*” that remain below the threshold of our awareness. He acknowledges that memories help to *show* that someone is the same as an earlier person, but he thinks that there must be a metaphysical foundation for that unity, namely a “monadic” substance that is defined by the complete set of perceptions it undergoes (Leibniz 1765: 2.27.9; GP V, 218).

## CONSCIOUSNESS AND PERSONAL IDENTITY

Butler worries that Locke's account is circular. For a mental state counts as a memory if and only if it represents a past state of the *person* whose state it is. A false memory counts as false because the *person* did not in fact have the experience in question. We need a notion of personal identity in order to secure the concept of memory (Butler 1736: 441). Indeed Locke himself seems to help himself to a memory-independent notion of personal identity in his account of the "Great Day," where God ensures that we are punished or rewarded only for the deeds that we in fact did; he reveals the "secrets of our hearts" in holding us responsible (Locke 1689: 2.27.26, 347; see also 2.27.22, 343–44). If our memories determined who we were in the past, there would be no secrets to be revealed; our not remembering them would mean we did not do them – someone else was their agent.

In order to interpret Locke as holding something other than a memory criterion for personal identity, we need a clearer grasp of how he understands consciousness, the crucial element of his definition of persons. His clearest statement on this topic is that "consciousness is the perception of what passes in a Man's own mind" (Locke 1689: 2.1.19, 115). Some interpreters have thus been tempted to equate consciousness with reflection, "that notice which the Mind takes of its own Operations, and the manner of them, by reason whereof there come to be *Ideas* of these Operations in the Understanding" (ibid.: 2.1.4, 105). Such an equation, however, runs afoul of several other Lockian tenets (see Kulstad 1991). First, Locke understands reflection to yield a kind of thought – the perception of ideas that represent our mental operations. And he takes all thought to be accompanied with consciousness. Just as all sensation is conscious, so also is all reflection. So if consciousness just is reflection, and all reflection is conscious, Locke would have introduced an infinite hierarchy of perceptions into the mind. Though Spinoza embraces such a conclusion (Spinoza 1677: 259), Locke recoils from attributing an infinite capacity to the mind. Second, Locke thinks that children rarely, if ever, reflect, because their attention is almost wholly fixed on the objects that they sense around them (Locke 1689: 2.1.24, 117–18). But children are conscious of all their sensations.

Accordingly, Locke cannot mean to equate consciousness with reflection. Instead, he seems to be developing an idea found in seventeenth-century Cartesians, such as Antoine Arnauld, who take consciousness to be a kind of self-awareness that is a *constituent* of every thought. In perceiving an idea of sensation or reflection, we would be conscious of ourselves as the ones so perceiving. We would be aware both of what the idea presents and of ourselves, the subjects of the awareness (Weinberg 2008; Coventry and Kriegel 2008). Thus, though Locke breaks with the Cartesian identification of the self with a thinking immaterial substance, he nonetheless retains a version of the Cartesian *cogito* argument: "In every Act of Sensation, Reasoning, or Thinking, we are conscious to our selves of our own Being [be it material or immaterial]; and, in this matter, come not short of the highest degree of *Certainty*" (Locke 1689: 4.9.3, 619). Consciousness is self-consciousness, where the self in question is just the person who is the subject of the consciousness. Moreover (and more controversially) this consciousness is not limited simply to our occurrent mental states. Locke allows that we "retain" or "extend" our thoughts in consciousness from one moment to the next (ibid.: 2.1.19, 115; 2.27.16, 340–41) without our having to remember them as prior thoughts. Thus our being *continuing* persons is manifested to us in all of our mental states.

DONALD AINSLIE AND OWEN WARE

Though Locke's view might escape the criticisms that were leveled at him for holding a memory criterion of personal identity, he faced second set of criticisms for his openness to the possibility of a material mind. Leibniz rejects this possibility in his famous "windmill" analogy from the *Monadology*:

Moreover, everyone must admit that *perception*, and everything that depends on it, is *inexplicable by mechanical principles*, by shapes and motions, that is. Imagine there were a machine which by its structure produced thought, feeling, and perception; we can imagine it as being enlarged while maintaining the same relative proportions, to the point where we could go inside it, as we would go into a mill. But if that were so when we went in we would find nothing but pieces which push one against another, and never anything to account for a perception. Therefore, we must look for it in the simple substance, and not in the composite, or in a machine.

(Leibniz 1720: §17)

His point seems to be that a material system, no matter how complex, can change only by having its parts rearranged, and no such rearrangement will suffice to be or produce thought. Monads, whether "bare monads" that lack consciousness, animal souls, or rational spirits, must be immaterial simple substances.

Samuel Clarke also rejects Locke's openness to a material mind, even while he accepts something like Locke's conception of consciousness (see Ahnert, Chapter 12, and Wolfe, Chapter 3, in this volume). It is the "Reflex Act by which I know that I think, and that my Thoughts and Actions are my own and not Another's" (Clarke 1706: 149). As in Locke, "the *Consciousness* that a Man has at one and the same time, is *One Consciousness*; and not a *Multitude of Consciousnesses*" (ibid.: 185). But Clarke holds that the unity of this consciousness means that it cannot be instantiated in a material system. If it were, the divisibility of matter would mean that each part of the system would bear a piece of the consciousness, undermining its unity. Where Locke thinks our fundamental ignorance of the real essence of matter means that we cannot rule out a material consciousness, Clarke argues that material systems can only bear qualities that are the same in kind as its parts. So unthinking matter cannot be combined to form a thinking mind.

Clarke develops this argument in an influential exchange with Anthony Collins, who defends a version of the Lockean position, in particular by suggesting that a system of matter (say, a rose or a brain) need not have qualities that are the same in kind as its parts (a petal is not a rose and a neuron is not a brain). Thus consciousness could occur in a material system without being a feature of its parts. Clarke presses Collins on the issue of how such a material system could preserve the unity of consciousness both at a time and through time, and Collins finally breaks with Locke's conception of consciousness as uniting our thoughts from moment to moment. Instead, Collins treats each state of consciousness as distinct from the next, so that "the individual Consciousness To-day, can neither in an individual or divisible Being be the individual Consciousness To-morrow: that Consciousness is a perfectly distinct Action in both Beings from the preceding Consciousness the Day before" (Clarke 1706: 236–37). He then embraces a memory criterion for personal identity, with all the attendant

## CONSCIOUSNESS AND PERSONAL IDENTITY

objections itemized above. Clarke concludes that Collins cannot account for such things as moral responsibility, and that only by recognizing that consciousness must inhere in an immaterial substance can we sustain our everyday understanding of ourselves (ibid.: 291–93).

### Hume: minds and persons

We have seen that the Lockean approach to persons fuses two different issues: a *metaphysical* one about what we are and a *moral* one about how we hold one another responsible in everyday life. Because, for Locke, every mental state includes an element of self-consciousness, we always manifest to ourselves what we are from the metaphysical point of view. And, because of the pervasiveness of this self-consciousness, it becomes fundamental for our moral self-understandings as well. Though Hume shares Locke's empiricism, he rejects Locke's account of self-consciousness and accordingly separates the metaphysical and moral issues about persons.

Hume's core metaphysical argument comes in his account of personal identity in the *Treatise of Human Nature*. Using his version of Lockean reflection, he "enters most intimately" into himself in order to "observe" his mental life, and he finds only a "bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement" (Hume 1739–40: 1.4.6.3; SBN 252). He does not find himself as the one having these perceptions. Hume thus breaks with Locke by suggesting that we are blind to our mental states, so that we are aware of *their content* without simultaneously being aware of *our having them*. We only become aware of our mental states when we reflectively observe them.

Hume's argument is perhaps not the strongest, for in the course of enunciating his point he seems to take for granted an understanding of himself as the one who engages in this inner observation: "For my part, when *I* enter most intimately into what *I* call *myself*, *I* always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. *I* never can catch myself at any time without a perception, and never can observe any thing but the perception" (Hume 1739–40: 1.4.6.3; SBN 252 – original emphasis on "myself" only). A Lockean could thus respond to Hume's view by arguing that, in reflecting on the mind, we both observe our mental operations, each of which is self-conscious, and at the same time are conscious of ourselves as the ones reflecting. Moreover, we recognize that it is *the same self* that both reflects and exerts itself in the various operations we observe by means of that reflection. Hume's response seems to be that in both philosophical observations of the mind and the vulgar observations of the world we remain blind to the perceptual mediation of our experiences. *Linguistically* we might invoke the first person, but the structure of the mind leaves us in the default position of taking ourselves to have direct access to what our perceptions reveal to us.

For Hume, then, it takes reflection for us to recognize our minds as bundles of perceptions. Most of us, however, observe our mental states only rarely and intermittently. So the view of the self as a bundle of perceptions – as a mind – cannot be the way we think of ourselves in everyday life. "We must have recourse to the most profound metaphysics to give a satisfactory" account of the mind; "in common life

DONALD AINSLIE AND OWEN WARE

'tis evident these ideas of self and person are never very fix'd nor determinate" (Hume 1739–40: 1.4.2.6; SBN 189–90). Thus Hume distinguishes between personal identity "as it regards our thought or imagination" and "as it regards our passions or the concern we take in ourselves" (ibid.: 1.4.6.5; SBN 253). The bundle view is a theory of only the former.

When we do observe our minds as bundles of perceptions, why do we believe that they are simple at a time – one individual mind – or identical through time? Hume answers this question by appealing to his associationist psychology. Though perceptions are, for Hume, all really distinct from one another, so that the bundle lacks any kind of intrinsic or "perfect" unity, it nonetheless has "imperfect" unity in the eyes of the observer (Hume 1739–40: 1.4.6.9; SBN 256). The causal and resemblance relations between the perceptions in the mind-bundle yield associations between an observer's ideas of these perceptions (ibid.: 1.4.6.16; SBN 260). Hume thinks that the association of ideas can often make us overlook any difference between the objects of the associated ideas. And thus the observer of the mind comes to feel as if the mind that she is observing is unified at a time and through time. He concludes that the mind is like

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.

(Hume 1739–40: 1.4.6.19; SBN 261)

Just as republics have relatively clear boundaries, Hume thinks that persons-as-minds usually have relatively clear boundaries. Even if someone has irremediably forgotten an event from his childhood the causal connections between it and the perceptions constituting the current mind-bundle mean that the forgotten experiences are still rightly attributed to the person who forgets them. But when you get to the Lockean puzzle cases, such as the prince and the cobbler, Hume thinks that there is no fact of the matter – no intrinsically unified self – that we can appeal to. "[A]ll the nice and subtle questions concerning personal identity can never possibly be decided, and are to be regarded rather as grammatical than as philosophical difficulties" (ibid.: 1.4.6.21; SBN 262).

Hume expresses second thoughts about his explanation of our beliefs about mental unity in the Appendix to the *Treatise*, published less than two years after the personal identity section first appeared. He declares that he has gotten himself into a "labyrinth" from which he sees no means of escape: there is an inconsistency between the claim that perceptions are "distinct existences" and the principle that the mind never discerns "real connexions among distinct existences" (Hume 1739–40: Appendix, SBN 636). Unfortunately Hume does not elaborate on why these two principles are inconsistent, and, given their centrality to a series of his other



## CONSCIOUSNESS AND PERSONAL IDENTITY

arguments in the *Treatise*, why their supposed inconsistency threatens only his account of personal identity.

Making sense of Hume's second thoughts on personal identity has become something of a cottage industry in the secondary literature, with various interpreters suggesting different options: the appeal to association presupposes an active principle beyond the bundle of perceptions (Robinson 1974; Passmore 1980: 82–83); the account Hume originally offers takes perceptions to come in bundles but he comes to realize he has no theory of the origin of such bundles (Pears 1990: 135–51; Haugeland 1998; Stroud 1977: 134–40); simultaneous non-spatial perceptions of, say, taste or smell cannot be localized to particular minds, in that they share all causal and resemblance relations with one another (Garrett 1981: 1997); Hume is mistaken in this second thoughts, having seemed to have lost the courage of his empiricist convictions (Beauchamp 1979; McIntyre 1979a; McIntyre 1979b; Swain 1991); and many others. Our preferred interpretation sees Hume's problem as arising from the relation between the reflecting mind that is observing its perceptions and the perceptions that are themselves observed (Ainslie 2001). Why do we believe that it is *one* mind that is both reflecting and constituted by the observed perceptions? His original treatment of the issue ignored the ideas by means of which we observe perceptions when we “enter most intimately” into ourselves. As we noted above, Hume holds that perceptions are normally invisible to us, so that we think only of their objects, not the mental states that afford our awareness of those objects. But, in giving his explanation of our belief in the unity of the perceptions that we observe when introspecting, Hume relies on the association of *ideas* of those perceptions (Hume 1739–40: 1.4.6.16; SBN 259–60). So he comes to recognize that the reflecting mind contains both observed perceptions and unobserved ideas of those perceptions. In the Appendix, Hume tells us that “consciousness is nothing but a reflected thought and perception” and that his second thoughts arise when he tries to explain the principles that unite perceptions in our “thought or consciousness” (Hume 1739–40: Appendix, SBN 635–36). We take this to mean that he has no way of explaining how we believe that the reflective mental states are part of the same mind as the one we observe by means of those states. Reflective perceptions are distinct existences lacking any real connections with the other perceptions in the bundle and yet we believe that they are part of one unified mind.

Though Hume came to have second thoughts about his treatment of personal identity as it regards the mind, he never backed away from his account of personal identity “as it regards our passions or the concern we take in ourselves” (Hume 1739–40: 1.4.6.5; SBN 253). The core of his treatment of this issue is found in the first two parts of Book 2 of the *Treatise*, where he addresses what he calls the “indirect” or person-oriented passions of pride, humility, love, and hatred. While much of his discussion focuses on the details of the mechanics of these passions, the “double relation of ideas and impressions” (*ibid.*: 2.1.5.5; SBN 286), he is best understood here as outlining a solution to the problem of why only some of a person's features are taken to make her who she is (Ainslie 1999). Of the many facts that are true of us, why are some “in a manner separated from us, and are never consider'd as connected with our being and existence” (Hume 1739–40: 2.1.8.9; SBN 302) while others define us as, say, Canadian, a philosopher, a member of this family, lazy, generous, this dog's owner, and the like? Hume answers that, because a defining feature of a person *matters* to us, we respond

DONALD AINSLIE AND OWEN WARE

to it positively or negatively. And, because the feature is connected in some respect with the person, the positive or negative reaction is “transfuse[d]” (ibid.: 2.1.5; SBN 290) into a positive reaction to the person overall. We come to have a conception of her or him as defined by that feature.

Hume puts various “limitations” (Hume 1739–40: 2.1.6; SBN 290–94) on the mechanism for the indirect passions, with the result that our conceptions of one another are not merely subjective, triggered by positive or negative reactions that we might or might not happen to feel. First, he thinks we only feel an indirect passion in response to a feature that is relatively rare. We react to people in a social context where we look for what sets them apart in some respect. Second, he thinks that others must feel the same way about the feature as we do. Because Hume takes sympathy – a tendency to feel the feelings of those who surround us – to be a fundamental feature of human nature, we cannot sustain a conception of someone that is not “seconded” by others (ibid.: 2.1.11.1; SBN 316). Note that as Norman Kemp Smith emphasizes, Hume’s explanation of sympathy invokes a special kind of ever-present self-consciousness that might seem to be in tension with his earlier rejection of the Lockean position:

*‘Tis evident, that the idea, or rather impression of ourselves is always intimately present with us, and that our consciousness gives us so lively a conception of our own person, that ‘tis not possible to imagine, that any thing can in this particular go beyond it. Whatever object, therefore, is related to ourselves must be conceived with a like vivacity of conception.*

(ibid.: 2.1.11.4; SBN 317 – emphasis added)

One of us has argued elsewhere that Hume’s uncertainty about this self-perception, indicated by the phrase “idea, or rather impression,” indicates that he means to treat this kind of self-consciousness in terms of the manner in which the different perceptions appear in the mind-bundle, rather than any discrete element of the bundle (Ainslie 2005). Thus he does not ultimately contradict his earlier anti-Lockean treatment of self-consciousness.

Finally, Hume thinks that “general rules” or “custom” influence the indirect passions. Just as his account of causation holds that, when we have repeatedly experienced objects conjointly, we come to expect them to continue their conjoint appearances, he also holds that, when most people tend to take a particular feature to be person defining, we suppose that others will as well, even if they do not show much interest in the feature in question. The result is that we construct social “ranks” into which people fall independently of their own feelings on the matter (Hume 1739–40: 2.1.6.8; SBN 293). The indirect passions are thus a mechanism whereby we are constructed as persons of particular kinds. Our emotional reactions to one another, spread through sympathy and regularized by general rules, create the social terrain we inhabit (ibid.: 2.1.6.9; SBN 293–94).

Hume focuses on four broad categories of features that tend to trigger the indirect passions and thus tend to define us as who we are: virtues and vices, bodily qualities (“beauty and deformity”), property and riches, and “external advantages and disadvantages” (Hume 1739–40: 2.1.7–10; SBN 294–316). This last is a catch-all category into which pretty much anything can fall so long as it is taken up into a society’s

## CONSCIOUSNESS AND PERSONAL IDENTITY

mutual feelings towards one another. He thus acknowledges trans-cultural commonalities in the human tendency to constitute persons in light of certain kinds of features as well as allowing for particular societies to have their own distinctive ways of making sense of themselves.

Hume's bundle theory of the metaphysics of persons had little influence in the eighteenth century. The *Treatise* fell "dead-born from the press" (Hume 1777: xxxiv), and the discussion of personal identity did not reappear in his later, more popular, works, perhaps because of the problem identified in its Appendix. Hume's account of the persons in everyday life and the constitution of rank through our passional responses to one another was more influential, notably in the later Scottish social philosophers several of whom in their own ways pursued Hume's project of understanding the social worlds that we create for ourselves: Adam Ferguson in the *Essay on the History of Civil Society* (1767), John Millar in *The Origin of the Distinction of Ranks* (1771), and Adam Smith in *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776).

### Rousseau: passions, persons, and politics

Hume's insight on the role of the passions in the constitution of the person was also developed independently by Rousseau in his *Discours sur l'origine et les fondemens de l'inégalité parmi les hommes* (*Discourse on the Origin and Foundations of Inequality among Men*) and *Émile, or On Education*, both of which were widely read and debated in the latter half of the century. Though Rousseau has little interest in the metaphysical question pioneered by Locke concerning the deep structure of persons, he, like Hume, thinks we come to be persons in a social sense only by the emotions we feel when contrasting ourselves with those around us (Dent 1998). But where Hume takes this feature of human nature in stride, Rousseau finds it deeply problematic. Our dependence on others' attitudes towards us means that our freedom is threatened unless we can remake society so as to render our mutual dependence on one another innocuous. Rousseau's *The Social Contract* (1762) sketches how society must change to allow for human freedom to flourish.

In the *Discourse*, Rousseau argues that distinctively human forms of social life emerge only through the operations of *amour-propre*, a kind of pride, self-love, or self-interest that he contrasts with *amour de soi*. Where the latter yields our concern for our personal welfare and survival, independently of the opinions of others, the former causes us to have a special interest in our status, determined in part through others' reactions to us (Rousseau 1755: n. O; OC III, n. 15, 219). Rousseau offers a speculative genealogy of human society, starting from a state of nature in which humans, governed solely by *amour de soi*, live freely but so primitively that they are hardly recognizable as human at all. He shows that it is only through the influence of *amour-propre* that such things as property, civilization, and morality, as well as hypocrisy, domination, and cruelty emerge: "[T]o this ardor to be talked about, to this furor to distinguish oneself, ... we owe what is best and worst among men: our virtues and our vices, our sciences and our errors, our conquerors and our philosophers" (ibid.: 184; OC III, 189). In *Émile* he traces a similar development, this time in

DONALD AINSLIE AND OWEN WARE

the individual, where the infant's *amour de soi* is usurped by *amour-propre* as he develops into a distinctively social person:

[A]s soon as *amour-propre* has developed, the relative *I* is constantly in play, and the young man never observes others without returning to himself and comparing himself with them. The issue, then, is to know what rank among his fellows he will put himself after having examined them.

(Rousseau 1762a: bk. 4, 243; OC IV, 534)

It is because of *amour-propre* that “sociable man [lives] always outside himself, knows how to live only in the opinion of others; and it is, so to speak, from their judgment alone that he draws the sentiment of his own existence” (Rousseau 1755: 187; OC III, 193).

Rousseau was often interpreted as wanting us to return to the state of nature, as if the effects of *amour-propre* could be extirpated from human nature. But it is clear that he recognizes *amour-propre* is the source of good as well as various evils. The challenge of *amour-propre* is that it leaves our very being as persons dependent on others' opinions of us. The fact that they are similarly dependent on us can lead to dangerous struggles, where each tries to control the terms that define us as who we are. Hypocrisy becomes widespread as people feel compelled to create a public facade that conforms to the social norms.

Rousseau's solution to the challenge posed by *amour-propre* is both individual and social (Neuhouser 2008). Education should be reformed so that we become men before we become citizens (Rousseau 1762a: bk. 1, 39; OC IV, 248) – the extent to which this point applies to women is less clear. Thus in *Émile* he argues that boys should be raised in isolation and in conformity with nature until their adolescence, when they can be introduced into society in such a way as to sympathize with others, to “see the same passions, the same sentiments in the hod-carrier and the illustrious man” (ibid.: bk 4, 225; OC IV, 509). Even if he has some talents that set him apart from others, he should learn to recognize their arbitrary nature: “[H]e is a man like others and subject to the same weaknesses” (ibid.: bk. 4, 245; OC IV, 537). Rousseau here anticipates Kant's claim that we possess a worth or dignity by virtue of our humanity, a worth that makes us genuine equals in society. But the need for a public identity remains, and Rousseau in *The Social Contract* argues that the state can provide it. Notably, every citizen should be an equal subject of the law (Rousseau 1762b: 1.6.10; OC III, 361–62). And every citizen should participate in the general will and thus collectively serve as sovereign, the source of law (ibid.: 2.4.5; OC III, 373). The law would then protect the freedom of all equally. This fundamental equality in place, the remaining differences among us that help to define us will be less dangerous.

### **Kant: apperception and autonomy**

While Rousseau and Hume think that the social, everyday self can be treated apart from the self-as-mind, Kant brings these two perspectives back together again. Like Locke, he emphasizes self-consciousness as fundamental to the self of thought, but he fills in the content for that self-consciousness in the practical realm, where we must

## CONSCIOUSNESS AND PERSONAL IDENTITY

view ourselves to be, like others, autonomous in our deeds. Kant can thus be seen as responding to Hume by unifying insights from Locke and Rousseau.

As we noted earlier, Leibniz accepts Locke's account of consciousness (what he calls "apperception") but rejects his claim that we are unable to draw further conclusions about the subject's constitution. Leibniz instead holds that the self is a monadic unity, a conclusion he draws on the basis of the unity of the conscious mind. Kant rejects this move both in Leibniz and in later eighteenth-century German philosophers who were influenced by him, such as Christian Wolff (1734), A. G. Baumgarten (1739), and Moses Mendelssohn (1767). While it is right to say that there cannot be an activity of thinking without a thinker, Kant argues we cannot infer from the "I think," as a formal feature of thought, to a thesis about what kind of thing "I" am: i.e. a simple substance distinct from my body. A thread common to these theories, which Kant groups under the heading "rational psychology," is the Cartesian assumption that we can infer real attributes of the soul (subsistence, simplicity, identity, etc.) from the structural or logical features of the "I" in "I think" (Kant 1781/1787: B407–9; 445–46). For Kant, we can speak of the self as a thing that thinks only if we keep our metaphysical status an open question.

Kant first raises the issue of self-consciousness in the "Transcendental Deduction of the Categories" from the *Critique of Pure Reason*. One of his main concerns in this section is to deny that the mind receives precombined representations of objects; instead, our sensibility yields the material for sensory thought while the mind is responsible for its combination. Cognition is thus a twofold process: we actively take up the data from sensibility and "synthesize" it into ordered representations. In defending this claim, Kant agrees with Hume that we do not perceive a fixed, abiding self in the content of sensible intuition. Consciousness of the self as given passively in inner sense lacks any necessary relation to a unified self, and the materials available to us through introspection reveal only a variety of mental states. Through empirical self-consciousness ("reflection," in Locke's and Hume's sense) we are nothing more than a "bundle of perceptions," none of which contain a self. As Kant writes: "The consciousness of oneself in accordance with the determinations of our state in internal perception is merely empirical, forever variable; it can provide no standing or abiding self in this stream of inner appearances, and is customarily called *inner sense*" (Kant 1781/1787: A107).

Unlike Hume, however, Kant does not think this supports skepticism about the intrinsic unity of the self. For Kant, the synthetic nature of cognition – our activity of combining representations of a manifold – presupposes such unity. This is not to say that in order to think or perceive we must *in fact* be self-conscious at a second-order level, by thinking that we think or perceiving that we perceive, but only that we *can* be aware of our unity through the relevant operations of cognition, operations that produce meaningful descriptions of objects or mental states precisely because they can be brought under first-personal ascription. Kant's point is that without this possibility of ascription, and so without relation to the "I think," representations of a manifold would be cognitively empty – "nothing" for us, as he says (Kant 1781/1787: B132) – because by definition they would not be *synthesized*. For this reason, Kant argues that we must have a special "consciousness" of our self-unity that is not passively received through inner sense, and this is what he calls "pure" apperception.

On this point his account is surprisingly close to Locke's theory of consciousness but, whereas Locke develops a psychological thesis about the thinking self, Kant's thesis is transcendental.

It is important to see that Kant's criticism of rational psychology, as well as his account of self-consciousness, follow from his division of reality into things-as-they-appear (*phenomena*) and things-in-themselves (*noumena*). Since our cognition of objects only concerns phenomena, rational psychology commits a category error: it takes structural features of self-consciousness for noumenal features of the soul. Kant instead shuts the door to theoretical knowledge of the self, in just the same way as his transcendental idealism rules out theoretical knowledge of other things in themselves. Yet there are crucial moments in Kant's philosophy where the door to self-knowledge reopens, if only slightly. In attempting to resolve the traditional debate between freedom and determinism (the "Third Antinomy" in the "Transcendental Dialectic"), Kant maintains that through pure apperception we have a special awareness of our self-activity, and this holds open our claim to freedom. Through intuition, we know ourselves only as phenomena, and as phenomena we must be subject to the same mechanistic laws that govern the relations between appearances. But since we cannot infer anything from appearances to things-in-themselves, it remains possible that we are, qua *noumena*, undetermined by the relations between appearances. It remains possible that we are free.

In making such claims, Kant comes dangerously close to a metaphysical thesis about the noumenal self, and thus violating his distinction between things-as-they-appear and things-in-themselves. At one point in the "Dialectic" he states: "The human being, who is otherwise acquainted with the whole of nature solely through sense, knows himself also through pure apperception, and indeed in actions and inner determinations which cannot be accounted at all among impressions of sense. ... We call these faculties understanding and reason" (Kant 1781/1787: A547/B575; see 1785: Ak IV, 452; 57). Kant is walking a fine line here. On the one hand, he officially rejects any theoretical knowledge of things-in-themselves, including ourselves as noumenal subjects. On the other hand, he seems to slip into talk that suggests we have some kind of access to our rational spontaneity. The fact that "I think" could accompany every representation reveals our synthetic activity and thus a unity of apperception that transcends the appearances. As Kant puts it, the representation of a manifold "is an act of spontaneity, i.e., it cannot be regarded as belonging to sensibility" (Kant 1781/1787: B132, 246), and later in the *Critique* he says we apperceive our capacity to think pure "Ideas" (as opposed to empirical concepts), the construction of which he assigns to reason rather than the understanding. In another text Kant goes so far as to say that the spontaneity of reason is unmixed with anything sensible – that in virtue of thinking ideas like God or freedom, or ideals like an ethical commonwealth or a Stoic sage, we are framing concepts that go "far beyond" anything sensibility can offer us through intuition (Kant 1785: Ak IV, 452; 57). In light of such comments, then, we might think Kant has succumbed to the very metaphysical temptations he accused his rationalist predecessors of.

One way of saving Kant from this verdict would be to appeal to his distinction between logical and real possibility. A thought that contains no contradiction presents us with a case of logical possibility; this is the status Kant assigns noumenal freedom in his resolution to the "Third Antinomy" (see Kant 1781/1787: A558/B586). By contrast,

## CONSCIOUSNESS AND PERSONAL IDENTITY

real possibility, or what Kant also calls objective validity, requires “more” than logical consistency, and Kant adds: “This ‘more,’ however, need not be sought in theoretical sources of cognition; it may also lie in practical ones” (ibid.: Bxxvi). This important, if obscure, suggestion will turn out to play a key role in Kant’s ethical theory during the 1780s and 1790s, for there he argues that we have grounds to ascribe freedom to ourselves, not intuitively, but practically, through our consciousness of the moral law. Put simply, Kant argues that our claim to freedom becomes objectively valid when we recognize ourselves under an obligation, for the actions that morality demands from us (actions we “ought” to perform in the eyes of reason) presuppose our freedom of will. The demand that I ought to act in such a way that I can will the universalization of my maxim (say, that I ought to help others, because doing so is a valid choice between persons and across situations), requires that I draw my motivation from the law itself. For Kant, moral action is unique in this way because it makes a general capacity of the will accessible to us: our capacity to determine ourselves, or what he calls “autonomy.” While this capacity underlies all of our choices, it is only visible in our encounter with the law. For “had not the moral law *already* been distinctly thought in our reason, we should never consider ourselves justified in *assuming* such a thing as freedom (even though it is not self-contradictory)” (Kant 1788: Ak V, 4n4; see also 1793: Ak VI, 26n; 74–75).

So the “more” that Kant says makes our claim to freedom objectively valid does not come from the sources of sensory experience. There is no corresponding intuition of freedom that would make it an object of cognition for us. From what he says elsewhere, it is likely Kant has in mind a kind of activity unique to pure practical reason. Whenever we are confronted with a decision, Kant says our reason will always give deliberative preference to itself, to its form, before it accepts any material incentive stemming from our sensible nature. Kant develops this argument in the “Analytic” of the *Critique of Practical Reason* (1788). Between an action that conforms to our inclinations and an action that conforms to the universality of a maxim, our reason will always adopt the latter as a standard of choice (Kant 1788: Ak V, 30). To be clear, Kant is not saying that we are always successful in following our reason over what we find desirable. He thinks we are at liberty to choose between the two at all times. But his point is that we must view ourselves as capable of acting on the basis of reason, and our evaluation of reasons must conform to a standard of justification. The universality of a maxim, the moral law, is the standard that reason upholds to us, whether or not we want it to. By “autonomy,” then, Kant does not mean anything like liberty of indifference. He means our capacity to evaluate actions under a rational norm. Our autonomy reflects a pure spontaneity in us because our evaluation of actions is not motivated by anything other than the moral law. In other words, our evaluation of actions is not driven by any impulse or inclination, even though, somewhat paradoxically, we are at liberty to endorse a maxim that aims toward satisfying the impulses and inclinations we have.

This appeal to the practical realm – to fill in the theoretically empty concept of ourselves as rational subjects – explains why Kant speaks of “persons” mostly in his ethical writings. As we have seen, the synthetic activity of cognition presupposes a unified subject, and Kant says that we have a special kind of consciousness of this activity that is revealed by the fact that we can attach the “I think” to any of our representations.

DONALD AINSLIE AND OWEN WARE

But this formal condition on our thought does not yet fulfill Kant's criterion of personhood. Personhood designates a capacity of the will to form maxims independently of external influences. While our consciousness of self-activity is revealed through the possibility of the "I think" relation, Kant constantly reminds us that we can never determine our spontaneity from a theoretical point of view. Theoretical cognition requires corresponding intuitive data, and we can never experience or intuit freedom of will. So he concludes that we can know ourselves as persons, not through the "I think" relation, but through the "I ought" relation. In this way, the door to self-knowledge reopens for Kant from a normative point of view: we know ourselves as persons when we are conscious of the moral law.

Like Rousseau, Kant believes this kind of self-knowledge is available to everyone, the philosopher and the common person alike. We do not need to engage in highly abstract thinking to recognize our moral autonomy; in fact, Kant believes speculative investigations always run the risk of going astray, so that philosophy might lead us to think we are unfree in our actions. Kant thanks Rousseau directly for helping him overcome his sense of intellectual superiority and to recognize the "dignity" of the common human being (1764–65: Ak XX, 44; 7). Indeed, Kant's account of the categorical imperative can be seen as transferring to the moral realm crucial elements of Rousseau's political theory of the general will discussed above, where each person serves as both the author and the subject of the law. So while Hume's skepticism about causation awoke Kant from his "dogmatic slumber," which led him to the system of transcendental idealism, it is fair to say that Rousseau helped Kant on the path that eventually led to his theory of what gives the human being dignity, namely, our autonomy.

### Conclusion

Kant's theories of apperception and autonomy had a profound impact on late eighteenth-century and nineteenth-century thought. Some of Kant's followers – notably K. L. Reinhold (1790) – started off as expositors of Kant's system, but their own attempts to explain the intricacies of the *Critique* opened up new conceptual territory that would later inform the work of J. G. Fichte (1794–95), F. W. J. Schelling (1795), and G. W. F. Hegel (1807) – three of the most influential thinkers in the German idealist tradition. One of the driving concerns of this tradition, especially in its early stages, was to articulate a deeper principle of self-consciousness that would unify Kant's views on the spontaneity of the cognitive thinker and the autonomy of the moral agent. For instance, in the *Grundlage der gesamten Wissenschaftslehre (Foundation of the Entire Science of Knowledge)* (1794–95), Fichte argued that we have immediate access to the active character of self-consciousness, such that we can understand the unity of the thinking self (theoretical reason) and the self of action (practical reason). A similar preoccupation with the unity of reason guided a number of important strands of thought in the early twentieth century, from Herman Cohen's neo-Kantianism (1871) to Martin Heidegger's brand of existentialist phenomenology as represented in his 1927 lectures (1988).

More recently, interpreters of Kant (Allison 1990, 1994; Korsgaard 1996; Wood 1999) and philosophers working in the area of self-knowledge and self-ascription



## CONSCIOUSNESS AND PERSONAL IDENTITY

(Burge 1998; Moran 2001; Bilgrami 2006) have tried to remove the air of mystery surrounding Kant's discussion of freedom by integrating his accounts of the spontaneity of action and the spontaneity of thought. If we can show that the kind of freedom Kant stipulates as a necessary condition for moral action does not differ from the kind of freedom we exercise in our cognitive activities, then we can make the transcendental foundation of Kant's ethical theory more palatable to modern readers. Along similar lines, philosophers working in the areas of normativity and action theory (Korsgaard 2008, 2009; Velleman 2000) have appealed to Kant in support of their view that we can ground norms of action or thought from "thinner" premises about what, at the most basic level, acting and thinking consist of.

While both projects contain important and exciting lines of inquiry, their connection to Kant seems textually thin. The same can be said about the idiosyncratic readings of Kant by the German idealists. For Kant, we are conscious of the spontaneity of thought through a kind of cognitive awareness called "pure apperception." Yet Kant denies we can ever know our absolute spontaneity, or freedom, in the theoretical realm. As agents, we know we are free only through our consciousness of the moral law. Only the moral law specifies actions that require freedom of will in the positive sense, i.e. our capacity for self-determination. There is nothing about the "I think" relation, or about our active representation of a manifold, that demands this kind of freedom from us. This is not to say that the spontaneity of thinking is actually limited, but only that our epistemic route to freedom does not have a theoretical source. Somewhat ironically, then, one of Kant's most important views has had the least impact on subsequent philosophy: his claim that there are inherent limitations to theoretical reason that only practical reason can overcome.

## References

---

- Ainslie, D. (1999) "Scepticism about Persons in Book II of Hume's *Treatise*." *Journal of the History of Philosophy* 37 (3): 469–92.
- . (2001) "Hume's Reflections on the Identity and Simplicity of Mind." *Philosophy and Phenomenological Research* 62 (3): 557–78.
- . (2005) "Sympathy and the Unity of Hume's Idea of Self." In C. Williams and J. Jenkins (ed.), *Persons and Passions*. South Bend: University of Notre Dame Press.
- Allison, H. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.
- . (1994) "Autonomy and Spontaneity in Kant's Conception of the Self." In *Idealism and Freedom: Essays on Kant's Practical Philosophy*. Cambridge: Cambridge University Press.
- Arnauld, A. (1683) *Des vraies et des fausses idées [On True and False Ideas: New Objections to Descartes' Meditations and Descartes' Replies]*. E. J. Kremer (trans.). Lewiston: Edwin Mellen Press, 1990.
- Ayers, M. R. (1981) "Mechanism, Superaddition, and the Proof of God's Existence in Locke's *Essay*." *Philosophical Review* 90: 210–51.
- Baumgarten, A. G. (1739) *Metaphysica*. In G. F. Meier (trans.) and D. Mirbach (ed.) (with notes by Johann August Eberhard) *Alexander Gottlieb Baumgartens Metaphysik*. 2nd ed. Jena, 1783.
- Beauchamp, T. L. (1979) "Self Inconsistency or Mere Self Perplexity." *Hume Studies* 5: 37–44.
- Bilgrami, A. (2006) *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press.
- Burge, T. (1998) "Reason and the First Person Concept." In C. Wright, B. Smith, and C. Macdonald (eds.) *Knowing Our Own Minds*. Oxford: Oxford University Press, 2000, 243–70.

DONALD AINSLIE AND OWEN WARE

- Butler, J. (1736) "Of Personal Identity." In *The Analogy of Religion*. 2nd ed. corr. London: John and Paul Knapton.
- Clarke, S. (1706) *A Letter to Mr. Dodwell*. 6th ed. London: James and John Knapton, 1731.
- Cohen, H. (1871) *Kants Theorie der Erfahrung*. Berlin: Dümmler.
- Coventry, A., and Kriegel, U. (2008) "Locke on Consciousness." *History of Philosophy Quarterly* 25: 221–42.
- Dent, N. J. H. (1998) "Rousseau on Amour-Propre." *Proceedings of the Aristotelian Society, Supplementary Volume* 72: 57–74.
- Downing, L. (1998) "The Status of Mechanism in Locke's *Essay*." *Philosophical Review* 107: 381–414.
- Ferguson, A. (1767) *An Essay on the History of Civil Society*. F. Oz-Salzberger (ed.). New York: Cambridge University Press, 1995.
- Fichte, J.G. (1794/95) *Grundlage der gesamten Wissenschaftslehre* [Foundations of the Entire Science of Knowledge] P. Heath and J. Lachs (trans. and ed.). Cambridge: Cambridge University Press, 1982.
- Garrett, D. (1981) "Hume's Self-Doubts about Personal Identity." *Philosophical Review* 90: 337–58.
- (1997) *Conviction and Commitment in Hume's Philosophy*. New York: Oxford University Press.
- Haugeland, J. (1998) "Hume on Personal Identity." In *Having Thought: Essays in the Metaphysics of Mind*. Cambridge, MA: Harvard University Press, 63–74.
- Hegel, G. W. F. (1807) *Phänomenologie des Geistes*. [Phenomenology of Spirit]. A. V. Miller (trans.). Oxford: Oxford University Press, 1977.
- Heidegger, M. (1988) *Die Grundprobleme der Phänomenologie* [The Basic Problems of Phenomenology]. Albert Hofstadter (trans.). Indianapolis: Indiana University Press.
- Hume, D. (1739–40) *A Treatise of Human Nature*. Oxford: Clarendon Press, 1978/2007.
- (1777) "My Own Life." In *Essays Moral, Political, and Literary*. E. F. Miller (ed.). Indianapolis: Liberty Press, 1985.
- Kant, I. (1764–65) *Bemerkungen in den "Beobachtungen über das Gefühl des Schönen und Erhabenen"* [Notes on the "Observations on the Feeling of the Beautiful and Sublime"]. Ak XX, 3–192. In P. Guyer (ed.) *Notes and Fragments*. Cambridge: Cambridge University Press, 2005.
- (1781/1787) *Kritik der reinen Vernunft* [The Critique of Pure Reason]. Ak III and IV, 5–252. P. Guyer and A. W. Wood (eds. and trans.). Cambridge: Cambridge University Press, 1998.
- (1785) *Grundlegung zur Metaphysik der Sitten* [Groundwork of the Metaphysics of Morals]. Ak IV, 387–463. M. Gregor (trans.). Cambridge: Cambridge University Press, 1997.
- (1788) *Kritik der praktischen Vernunft* [Critique of Practical Reason]. Ak V. In M. J. Gregor (ed. and trans.) *Practical Philosophy*. Cambridge: Cambridge University Press, 1996, 1–163.
- (1793). *Die Religion innerhalb der Grenzen der blossen Vernunft* [Religion Within the Boundaries of Mere Reason]. Ak VI, 3–202. G. Giovanni (trans.). Cambridge: Cambridge University Press, 1996.
- Korsgaard, C. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- (2008) *The Constitution of Agency*. Oxford: Oxford University Press.
- (2009) *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Kulstad, M. (1991) *Leibniz on Apperception, Consciousness and Reflection*. Munich: Philosophia.
- Leibniz, G. W. (1720) *Monadology*. In R. S. Woolhouse and R. Francks (eds.) *Philosophical Texts*. Oxford: Oxford University Press: 1999.
- (1765) *Nouveaux essais sur l'entendement humain* [New Essays on Human Understanding]. P. Remnant and J. Bennett (trans.). Cambridge: Cambridge University Press, 1981.
- Locke, J. (1689) *An Essay concerning Human Understanding*. P. Nidditch (ed.). Oxford: Oxford University Press, 1975.
- (1742) *Essai philosophique concernant l'entendement humain*. P. Coste (trans.) and G. J. D. Moyal (ed.). Paris: Vrin, 2004.

CONSCIOUSNESS AND PERSONAL IDENTITY

- McCann, E. (1994) "Locke's Philosophy of Body." In V. Chappell (ed.) *The Cambridge Companion to Locke*. Cambridge: Cambridge University Press.
- McIntyre, J. L. (1979a) "Further Remarks on the Consistency of Hume's Account of the Self." *Hume Studies* 5: 55–61.
- . (1979b) "Is Hume Self-Consistent?" In D. F. Norton, N. Capaldi, and W. L. Robinson (eds.) *McGill Hume Studies*. San Diego: Austin Hill.
- Mendelssohn, M. (1767) *Phädon oder über die Unsterblichkeit der Seele*. [Phaedon or on the immortality of the soul]. In F. Bamberger and L. Strauss (eds.) *Gesammelte Schriften Jubiläumsausgabe*, v. 3.1: *Schriften zur Philosophie und Ästhetik*. Berlin: 1972.
- Millar, J. (1771) *The Origin of the Distinction of Ranks*. A. Garrett (ed.). Indianapolis: Liberty Fund, 2006.
- Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- Neuhouser, F. (2008) *Rousseau's Theodicy of Self-Love: Evil, Rationality, and the Drive for Recognition*. Oxford: Oxford University Press.
- Passmore, J. A. (1980) *Hume's Intentions*. 3rd ed. London: Duckworth.
- Pears, D. (1990) *Hume's System*. Oxford: Oxford University Press.
- Reid, T. (1785) *Essay on the Intellectual Powers of Man*. D. R. Brookes (ed.). University Park: Pennsylvania State University Press, 2002.
- Reinhold K. L. (1790) *Briefe über die Kantische Philosophie*. Leipzig: Georg Joachim Göschen. English translation of the first eight letters in the versions originally published in *Der Teutsche Merkur*, supplemented by "the major additions in the 1790 edition." In J. Hebbeler (trans.) and Karl Ameriks (ed.) *Letters on the Kantian Philosophy*. Cambridge: Cambridge University Press, 2005.
- Robinson, W. (1974) "Hume on Personal Identity." *Journal of the History of Philosophy* 12: 181–93.
- Rousseau, J.-J. (1755) *Discours sur l'origine et les fondemens de l'inégalité parmi les hommes*. In V. Gourevitch (trans. and ed.). "The Discourses" and Other Early Political Writings. Cambridge: Cambridge University Press, 1997.
- . (1762a) *Émile ou De l'Éducation* [Emile: or, On Education]. A. Bloom (trans.). New York: Basic Books, 1979.
- . (1762b) *Du contrat social*. In V. Gourevitch (trans. and ed.) "The Social Contract" and Other Later Political Writings. Cambridge: Cambridge University Press, 1997.
- Rozmond, M., and Yaffe, G. (2004) "Peach Trees, Gravity, and God: Mechanism in Locke." *British Journal for the History of Philosophy* 12: 387–412.
- Schelling, F. W. J. (1795) *Vom Ich als Prinzip der Philosophie oder über das Unbedingte im menschlichen Wissen* [Of the I as the principle of philosophy or on the unconditional in human knowledge]. In L. F. Marti (trans.) *The Unconditional in Human Knowledge: Four early essays 1794–1796*. Lewisburg: Bucknell University Press, 1980.
- Smith, A. (1776) *An Inquiry into the Nature and causes of the Wealth of Nations*. A. S. Skinner (ed.). Oxford: Oxford University Press, 1976.
- Spinoza, B. (1677) *Ethics*. In S. Shirley (trans.) and M. L. Morgan (ed.) *Spinoza: Complete Works*. Indianapolis: Hackett, 2002.
- Stroud, B. (1977) *Hume*. London: Routledge.
- Swain, C. (1991) "Being Sure of Oneself: Hume on Personal Identity." *Hume Studies* 12: 107–24.
- Velleman, D. (2000) *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Weinberg, S. (2008) "The Coherence of Consciousness in Locke's Essay." *History of Philosophy Quarterly* 25 (1): 21–40.
- Wilson, M. (1979) "Supperadded Properties: The Limits of Mechanism in Locke." *American Philosophical Quarterly* 16: 143–50.

DONALD AINSLIE AND OWEN WARE

- Wolff, C. (1734) *Psychologia rationalis*. In J. École et al. (eds.) *Gesammelte Werke*. Hildesheim: Olms, 1962.
- Wood, A. (1999) *Kant's Ethical Thought*. Cambridge: Cambridge University Press.

### Further reading

---

D. Ainslie, "Sympathy and the Unity of Hume's Idea of Self," in C. Williams and J. Jenkins (eds.) *Persons and Passions* (South Bend, IN: University of Notre Dame Press, 2005), 143–73;

H. Allison, "Autonomy and Spontaneity in Kant's Conception of the Self," in *Idealism and Freedom: Essays on Kant's Practical Philosophy* (Cambridge: Cambridge University Press, 1994), 129–42;

A. LoLordo, *Locke's Moral Man* (Oxford: Oxford University Press, 2012);

P. Keller, *Kant and the Demands of Self-Consciousness* (Cambridge: Cambridge University Press, 1998);

M. Kulstad, *Leibniz on Apperception, Consciousness, and Reflection* (Munich: Philosophia, 1991);

U. Thiel, *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume* (Oxford: Oxford University Press, 2011);

J. Yolton, *Thinking Matter: Materialism in Eighteenth-Century Britain* (Minneapolis: University of Minnesota Press, 1984).