The Self: who or what are we?

(Paper for Kingston Philosophy Café online session on 27 January 2021)

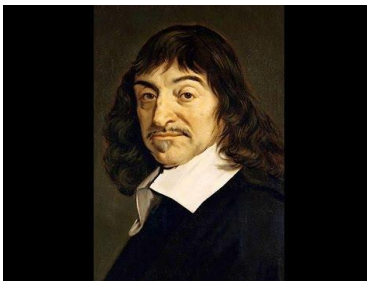
Any coherent account of the 'self' must encompass both its mental and physical aspects.

1. Distinguishing different selves or persons is something we do pre-reflectively and generally without problem. Our everyday use of personal pronouns – 'I', 'me', 'you', etc. – might appear uncontentious and conceptually unchallenging. However, when we try to pin down just what is denoted by these terms, things become less clear. Whilst the *sense* of self which we naturally possess includes awareness of *embodimen*t and *relative spatial position*, we do not *equate* ourselves with our bodies. Crucially, we are aware of ourselves as *thinking* and *feeling* beings. Conceptual issues about the 'self' stem primarily from this apparent duality – mental and physical – of its properties and how they might possibly interrelate.

Substance dualism ascribes the mental and physical attributes of the self to distinct substances.

2. *Substance dualism* – a view of reality propounded most famously by René Descartes (1596-1650) but shared over the centuries by many other philosophers and probably, in some form or another, by most people – ascribes mental and physical phenomena to distinct substances with different modes of existence. According to Descartes:

> "If we perceive the presence of some attribute, we can infer that there must also be present an existing thing or substance to which it may be attributed... each substance has one principal property which constitutes its nature and essence and to which all its other properties are referred. Thus extension... constitutes the nature of corporeal substance; and thought constitutes the nature of thinking substance"[PP 1.52-53].



In this way, Descartes identifies two types of substance: *extended* substance (body/matter) and *thinking* substance (spirit/mind). Ascribing to different substances the mental and physical properties of the self, however, raises obvious – but not obviously answerable – questions about how they could possibly *co-exist* and *co-function* within a supposed *unitary* being.

The relationship between minds and bodies, conceived as two distinct substances, is problematic.

3. John Locke (1632-1704) adds solidity, separability and moveability to the properties of corporeal substance but, for the most part, goes along with Descartes' dualist distinction. He recognises, nevertheless, some of the complications which arise regarding the relationship between putative 'spirits' and their bodies. He is forced, for example, to consider position and motion as properties of spirits as much as of their associated bodies:

> "Spirits as well as bodies cannot operate but where they are ... Everyone finds in himself that his soul can think, will and operate on his body in the place where that is; but cannot operate on a body or in a place an hundred miles distant from it. Nobody can imagine that his soul can think or move an object at Oxford whilst he is in London; and cannot but know that, being united to his body, it constantly changes place all the whole journey between Oxford and London, as the coach or horse does that carries him" [ECHU 2.23.19-20].

***Immaterialist monism* purports to avoid the problem by denying the existence of *physical* substance.**

4. Our perceptual/mental experience is affected not only by the location, but also the *condition*, of our bodies/brains – most obviously by the acuity of our senses, whether we are awake or asleep and the impact of drugs, injury or disease upon brain functioning. Again, the question arises as to how a spirit/soul

comprising 'thinking substance' could be affected in any way by a body/brain comprising 'corporeal substance', each supposedly having *in*dependent, not *inter*dependent, modes of existence. Is the answer to jettison one of the substances and adopt a *monist* view of reality? George Berkeley (1685-1753) does just this and opts to sacrifice corporeal substance. He declares it "evident there is no other Substance than Spirit, or *that which perceives*" [PHK 7], that "a spirit is one simple, undivided, active being" [PHK 27] which

is "indivisible, incorporeal, unextended" [PHK 141], and that "this perceiving, active being is what I call *mind, spirit, soul* or *myself*" [PHK 2]. He explains that by these words he does "not denote any one of my ideas, but a thing entirely distinct from them, wherein they exist, or, which is the same thing, whereby they are perceived – for the existence of an idea consists in being perceived" [PHK 2]. Accepting that "there can be no idea formed of a soul or spirit; for, all ideas whatever, being passive and inert, cannot represent unto us, by way of image or likeness, that which acts", he can offer only the vague suggestion that "we have some notion of soul, spirit and the operations of the mind; such as willing, loving, hating – inasmuch as we know or understand the meaning of these words" [PHK 27].

Our inability to *choose* our sensory 'ideas' demands an explanation of their *source*.

5. An obvious problem (amongst many) with Berkeley's formulation is that if human selves consist of nothing else but 'Spirit' – the *only* substance which exists – what is the status of the sensory and other ideas which they experience? According to Berkeley, ideas are not themselves a type of substance and have no independent existence of their own. A critical problem for Berkeley and other proponents of immaterialism is to explain what determines our sensory experience if this is not of *something* which has a separate and independent existence. Berkeley recognises that we do not *choose* such experience and, having ruled out the existence of material substance, can attribute it only to the implanting of sensory ideas ('sensations') in our minds by a supreme spirit (God). "When in broad daylight I open my eyes, it is not in my power to choose whether I shall see or no, or to determine what particular objects shall present themselves to my view; and so likewise as to the hearing and other senses, the ideas imprinted on them are not creatures of my will. There is therefore some other Will or Spirit that produces them" [PHK 29].

If they are mere combinations of *passive* ideas, what is the *purpose* of our bodies and body parts?

6. Berkeley argues [PHK3] that we *combine/blend together* God-implanted sensory ideas to compose *everything* we perceive – which includes, of course, our own bodies and body parts. Thus the eyes which Berkeley says he opens in broad daylight are as much 'combinations of ideas' as anything else and equally *functionless*. If everything we perceive, including all life forms, comprise *passive* ideas – which "being ideas, have nothing powerful or operative about them, nor have any necessary connection with the effects ascribed to them" [PHK60] – what is the *purpose* of the structures they display? Berkeley acknowledges the problem but provides no answer, only the obscure suggestion that "though the fabrication of all those parts and organs be not absolutely necessary to the producing *any* effect, yet it is necessary to the producing of things *in a constant regular way according to the laws of nature*" [PHK 62].

How can *disembodied* human 'spirits' *intercommunicate* and act as *moral agents*?

7. Fatal to Berkeley's immaterialist doctrine is its inability to account for *human agency* – and thus for *moral responsibility*. If the entirety of our sensory experience arises from God-implanted 'ideas', how do we operate as agents who make choices and whose actions affect the sensory experience of others? According to Berkeley, we are *disembodied* spirits lacking dimension or spatial position and what we perceive as our bodies, including our limbs and vocal chords, are just combinations of *passive* ideas. So how can human spirits interact and intercommunicate? This is possible, in Berkeley's ontology, only if God acts as an intermediary and implants relevant sets of sensory ideas in their minds. "For it is evident that in

affecting other persons the will of man has no other object than barely the motion of the limbs of his body;[1] but that such motion should be attended by, or excite any idea in the mind of another, depends wholly on the will of the Creator. He alone it is who ... maintains that intercourse between spirits whereby they are able to perceive the existence of each other" [PHK 147]. The problem is that people may desire *contradictory* things and those who prevail can do *bad* as well good things to others. As the intermediary between human spirits, God must choose *which* of their opposing intentions to realise in the form of implanted ideas and thus becomes *complicit* in all the bad things which undoubtedly happen in the world – a point made by Hylas, one of the two protagonists in Berkeley's *Three Dialogues between Hylas and Philonous* (1713): "You are not aware, Philonous, that in making God the immediate author of all the motions in nature, you make him the author of murder, sacrilege, adultery, and the like heinous crimes" [DHP3 H25]. Philonous' reply – that sin lies in the intention not the act – fails to address the point.

An 'expanded notion of the physical' might provide a coherent alternative to dualist and immaterialist approaches with all their inherent problems. Locke recognises such a possibility.

8. If both substance dualism and immaterialist monism run into irresolvable conceptual difficulties, is there a credible alternative? Might an approach more consonant with our life experience be found by rejecting as mistaken the supposition "that the common-sense distinction between mental states naively construed and physical states naively construed is an expression of some deep metaphysical distinction" and by accepting instead "an expanded notion of the physical to allow for its intrinsic, subjective mental component" (Searle, 2004). Whilst unable to abandon belief in a divine and omnipotent Spirit, Locke does at one point countenance such a possibility: “We have the ideas of *matter* and *thinking* but possibly shall never be able to know whether any mere material being thinks or no; it being impossible for us, by the contemplation of our own ideas, without revelation, to discover whether Omnipotency has not given to some systems of matter fitly disposed, a power to perceive and think, or else joined and fixed to matter so disposed a thinking immaterial substance: it being, in respect of our notions, not much more remote from our comprehension to conceive that God can, if he pleases, super-add to matter a faculty of thinking, than that he should super-add it to another substance with the faculty of thinking.” [ECHU 4.3.6]

Consciousness, including *self*-consciousness, can be viewed as an *emergent* property of highly complex biological/cognitive systems – notably *humans*.

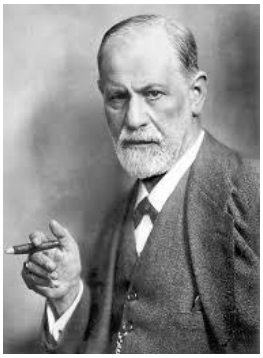
9. Locke's suggestion that 'some systems of matter', if 'fitly disposed', might display 'a power to perceive and think' is consistent with the view that consciousness is an *emergent property* of biological *systems* (most obviously humans) which have evolved *complexly structured brains*. If true, this has crucial implications for the existential status of the 'self'. Selves become *constructs*, along with a host of others, of the cognitive systems (minds) realised within such brains. Substantial consistency in what is experienced is provided by the ability of brains to *store*, ready for activation, the attributes of those constructs. Variations in exactly what is activated, however, may result in corresponding variations in the selves experienced at different times and under different circumstances. This squares with the varying images we may have of ourselves and the varying ways in which we may behave/react depending upon our current mood and physical/social setting. The self/person experienced internally by ourselves and externally by others is subject not only to gradual change as we go through life but also to *sudden* change as we react to different circumstances, an extreme example being the ability of some of us, if sufficiently provoked, to flip in a matter of seconds from a rational, benign and peace-loving person (at least in our own estimation) to an out-of-control, angry and vengeful monster – as evidenced by the phenomenon of road rage. *Sustained* inconsistency of behaviour is also evident – a frightening example being the loving parents who were amongst those complicit in the murder of other people's children in Nazi extermination camps (see Appendix A). *None* of this squares with the notion of the self as an immaterial spirit comprising, in Berkeley's words, “one simple, undivided, active being".

[1] In suggesting that spirits might impart motion to their limbs, Berkeley *contradicts* his own immaterialist doctrine. In his ontology, limbs are as much combinations of passive God-implanted sensory ideas as anything else.

Variations of self-image and behaviour are consistent with 'selves' as *products* of cognitive systems.

10. The images we conjure up of ourselves and of others are *composites* of mental and physical attributes – such as appearance, constitution, temperament, patterns of belief and behaviour, social status and moral worth – and may vary significantly in terms of content and degree of integration. As stated by Argyle (1988): "The more integrated the self-image, the more consistent a person's behaviour will be: one effect of the self-image on behaviour is the suppression of behaviour that is out of line. This 'consistency' may take various forms, depending on whether the self-image is based on the attributes of some person, or on a set of ethical or ideological rules of conduct, or on an occupational or social-class role." There can be wide differences between our self-images and the images which others have of us – hence the corrective potential of having the gift, if it were possible, "to see ourselves as others see us"[2]. Whilst we may feel we know ourselves far better than anyone else ever could, there is, as argued by Ryle (1949), "no contradiction in asserting that someone might fail to recognise his frame of mind for what it is" and that people might "deceive themselves about their own motives". The reality of self-deception is inconsistent with the notion of the self as a unitary being which enjoys privileged and indubitable knowledge of its own nature. It *is* consistent, on the other hand, with the self as the intermittent and variable product of a *semi*-integrated cognitive system operating at *more than one level of consciousness* and capable of generating, to some extent at least, different experiences of selfhood on different occasions – an extreme and pathological example being dissociative identity disorder (previously called *multiple person* disorder).

Freud provides a tripartite model of the human mind and distinguishes three levels of awareness.



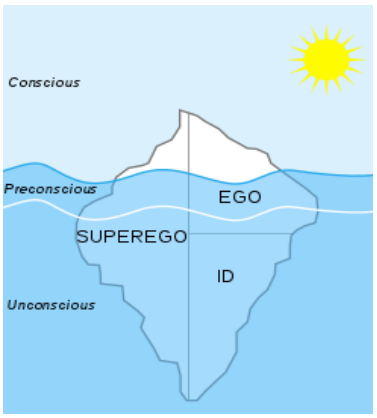
11. Sigmund Freud (1856-1939) represents the human mind as split into three parts which, through their *interaction*, determine human behaviour:

- the *id* comprising instinctive urges and desires, including sexual and aggressive drives;
- the *ego* which mediates between the competing demands of the *id* and the *superego* and seeks, by *reasoning*, to reconcile both with manifest *reality*;
- the *superego* which acts as a self-critical *conscience*, applying social norms learned primarily from parents and teachers.

Although Freud's tripartite model is viewed in modern psychology as simplistic and lacking clear evidential support, it is consistent at least with the experience we commonly have, when deciding how to act, of being pulled in different directions by competing thoughts and feelings and of conflicts between what we instinctively *want* to do and what we judge we *ought* to do. The more variation at different times and under different circumstances in the *relative strength* of competing influences, the less consistent will be our behaviour. Whilst not the first to do so, Freud recognises the influence upon our behaviour of factors of which we may be unaware and divides mental events into:

- the *conscious* (those of which we are aware);
- the *preconscious* (those of which we are in process of becoming aware);
- the *unconscious* (those of which we remain unaware).

For Freud, mental events associated with the *id* are entirely 'submerged' at the level of the unconscious whilst those associated with the *ego* and the *superego* are to be found at all three levels – an iceberg analogy sometimes being used to illustrate the supposed arrangement.



[2] *To a Louse, On Seeing One on a Lady's Bonnet at Church* (1786) by Robert Burns (1759-96)

Modern neuroscience substantiates the impact of unconscious processes on human behaviour.

12. There is no correspondence, it is important to emphasise, between the varied and highly complex structures of the *brain* and Freud's tripartite division of the *mind*. Modern neuroscience has revealed much about the brain's component parts, including their individual functions and how they work together as a system to generate the diverse experiences we associate with *consciousness* – described by Gibb (2012) as "an emergent property of the brain as a whole, a natural consequence of millions of neurons processing information in parallel". Whilst not validating Freud's model of the human mind, modern neuroscience has substantiated the existence of processes which profoundly affect our behaviour but of which we are unconscious or, at best, only dimly aware. "In perpetual homage to Sigmund Freud", says Swaab (2014), "our behaviour is for a very great part steered by unconscious processes. A hundred years later we have returned to the subconscious, but this time without the Freudian vision of repressed, infantile, sexual, and aggressive urges and other dubious claims".

The conscious outcomes of unconscious processes are amenable to examination and appraisal.

13. If unconscious processes steer much of our behaviour, how far can we be deemed free to choose our actions and thus to be responsible for them? Many of such processes, of course, simply allow us to interact with our environment (e.g. cycle on a busy road) without having to be conscious of each and every action involved. Of principal concern are the processes affecting our *moral* choices – i.e. those about what we *should* and *shouldn't* do – and, in particular, the relative influence of *rational* and *emotional* factors. Thoughts and feelings, although by their nature something of which we are conscious, may be triggered and conditioned by internal or external stimuli of which, at the time at least, we are unaware. Arguably, however, those thoughts and feelings, by virtue of entering our consciousness and regardless of whatever might have stimulated them, become open to examination and appraisal and *themselves* the stimuli for, and potential objects of, *other* thoughts and feelings. One moment, for example, another road user's bad driving may be the object of my anger. The next moment that anger itself may become the object of my attention. Self-examination and self-criticism are common features of our everyday experience and crucial to redirecting our behaviour. Through them, retrospectively at least, we *can* gain awareness of factors which subconsciously influence our thoughts, feelings and behaviour. Our attention might also be drawn to them by the observations of others – including those of psychiatrists and neuroscientists!

Reason alone is insufficient for moral choice. This requires an *emotional* response to rationally identified possibilities.

14. In the determination of moral choice, reason and emotion are often seen as *competing* influences. There is evidence, however, that a *combination* of the two is essential for effective decision-making. The insights of neuroscience into the workings of the human brain/mind have been gained in large measure from the study of people in whom those workings have been *impaired* by, for example, injury or disease. One such person – described by Gibb (2012) as a "real-life Mr Spock" (the hyper-logical character in the TV/film series *Star Trek*) – had a brain tumour successfully removed but not before it had destroyed parts of his frontal cortex, an area of the brain associated with reasoning and decision-making. The damage, however, did not seem to affect his reasoning powers, including his capacity to identify and explore possible scenarios and solutions to problems – only his ability to *choose between them*. What appeared to have been destroyed was a cortical connection enabling emotional responses to rationally identified possibilities. Without this he was trapped in an entirely rational world, his mind overloaded with possibilities but unable to decide which to pursue. Gibb concludes: "In the undamaged brain, as the mind moves through a number of possible choices, it is the emotions that give the thumbs up or down, by fleetingly providing an insight into how the consequences of a specific choice would make us feel. However much it goes against our conception of ourselves as rational creatures, the role of the emotions in decision-making cannot be overstated".

The relationship between reason and emotion is essentially *two-way*. The rationality of our emotional impulses is open to challenge, as a result of which they may be moderated.

15. In portraying emotions as giving 'the thumbs up or down' to possible choices, Gibb seems to imply that they *predominate* over reason in determining our behaviour. David Hume (1711-76) certainly ascribes to them a predominant role when he asserts: "Reason, is, and ought only to be, the slave of the passions, and can never pretend to any other office than to serve and obey them" [THN 2.3.3]. Arguably, however, the interaction between reason and emotion is much more complex and essentially *two-way*. Our emotional impulses – which can reflect all sorts of desires, urges, prejudices and presumptions – are *themselves* open to examination, appraisal and correction. Emotions, (unlike, for example, toothaches) are *intentional* phenomena, the *rationality* of which may be judged on the basis of *evidence*. We may judge, for example, whether a fear is justified by the existence of a genuine danger or whether it constitutes a *phobia*. Thus xenophobia, homophobia and the like can be challenged by demonstrating through *rational argument* and *evidence* that the hatreds involved are based on *untruths*. We are capable of recognising irrational impulses in ourselves as much as in others and it is by consciously seeking to moderate, if not eliminate, their influence that we foster our own moral development. Acting morally involves the application of clear and consistent criteria for behaviour rather than the arbitrary exercise of gut feeling. Unless we are prepared to accept the objective existence of 'imperatives' which are either self-revelatory or revealed to us by shamans claiming privileged insight into a hidden reality, the identification and refinement of such criteria – which are always open to revision in the light of experience – require our on-going and individual application of both emotional sensibility and reason to real and imagined situations. Through such a process we can each develop a moral outlook which is genuinely our *own* and which we are prepared, if challenged, to defend. Thus, whilst we may be strongly influenced by the teachings and examples (bad as well as good) of parents and others, especially in our early years, we need not be *enslaved* by them.

David Hume identifies himself as "nothing but a bundle or collection of different perceptions". But can a bundle of perceptions perceive itself to be so?

16. As evidenced already in this paper, it is difficult, if not impossible, to discuss the self without pre-reflective use of words, especially personal pronouns, referring to the very thing under consideration. In reporting his inability to 'catch' himself, David Hume uses the pronoun 'I' to denote an entity (himself) which, when it reflects upon itself, can perceive only perceptions.[3]

"I can never catch myself at any time without a perception, and never can observe anything but the perception. When my perceptions are removed for any time, as by sound sleep, so long am I insensible of *myself*, and may truly be said not to exist. And were all my perceptions removed by death, and could I neither think, nor feel, nor see, nor love, nor hate, after the dissolution of my body, I should be entirely annihilated, nor do I conceive what is further requisite to make me a total nonentity." [THN 1.4.6]

"When I turn my reflection on *myself*, I never can perceive this *self* without some one or more perceptions; nor can I ever perceive anything but the perceptions. It is the composition of these, therefore, which forms the self". [THN A]

Hume concludes that a self is "nothing but a bundle or collection of different perceptions, which succeed each other with inconceivable rapidity, and are in a perpetual state of flux and movement" [THN 1.4.6]. If so, the '*I*' to which he refers is not distinct from but *comprises* just such a bundle. He thus asserts, in effect,

[3] Hume uses the word 'perception' in much the same way as Locke and Berkeley use the word 'idea' – i.e. as a *generic* term to cover *all* forms of sensory and cognitive experience including sights, sounds, touches, tastes, smells, pleasures, pains, hopes, fears, desires, intentions, memories, imaginings and everything that might count as thoughts or feelings. Hume subdivides perceptions into simple or complex *impressions* and *ideas*. The problem in all cases is to identify the *units* involved i.e. the *temporal subdivisions* of sensory/cognitive experience constituting *individual* perceptions.

that whenever the bundle constituting himself introspects, all it can observe is one or more of its own perceptions. Is this intelligible? Can a bundle of perceptions be aware of itself as a bundle of perceptions?

If perceptions are *fleeting* 'distinct existences', how can they form into 'bundles'? Hume, nevertheless, ascribes to such bundles a faculty of memory by which they "discover" their "personal identity".

17. Hume clearly rejects the notion of the self as an entity existing independently of any perceptual experience or associated body and so unaffected by their cessation/dissolution. He less certain, however, about his conception of the self as a 'bundle of perceptions', recognising that it requires an explanation of "the principles that unite our successive perceptions in our thought or consciousness". He confesses himself unable to render consistent two principles, neither of which he feels able to renounce: "viz. *that all our distinct perceptions are distinct existences*, and *that the mind never perceives any real connection amongst distinct existences*" [THN A]. His difficulty arises not just from the elusiveness of a uniting principle but also from his *atomistic* model (acquired from Locke) of the perceptions to be united.[4] With regard to a uniting principle, the best he can suggest is that by finding causative relationships between remembered perceptions and extending these to fill gaps in our memory, we discover our identity as persons. "Memory does not so much *produce* as *discover* personal identity by showing us the relation of cause and effect among our different perceptions... Having once acquired this notion of causation from the memory, we can extend the same chain of causes, and consequently the identity of our persons beyond our memory, and can comprehend times, and circumstances, and actions, which we have entirely forgot, but suppose in general to have existed" [THN 1.4.6]. Once again, Hume uses personal pronouns – in this case, '*we*' and '*us*' – to denote beings which, by his own account, are *just* bundles of perceptions. He thus ascribes to such bundles a faculty of memory, a notion of causation and the ability, by extending causative chains, to discover their own 'personal identity'. But how can bundles of disparate perceptions function as *unitary beings* in this way? If there is no "*real connection* amongst distinct perceptions", moreover, what ties any particular perception to any particular bundle? Crucially, if perceptions are in a "*perpetual state of flux*" and "*succeed each other with inconceivable rapidity*", they exist only *momentarily* and cannot then *co-exist* in the form of 'bundles' comprising 'selves'.

Memories are imperfect *simulations* of perceptual experiences, key features of which must have been encoded and stored in such a way as to allow for their subsequent retrieval and decoding.

18. Although all perceptions are fleeting, some may be *simulated* at a later stage by *memories* – defined by Hume, along with the products of the imagination, as "kinds of ideas" [THN 1.1.3]. So-defined, memories are *themselves* perceptions.[5] They comprise *remembrances* – i.e. *acts* of recall – which, as they occur, are as fleeting as the perceptual experiences they simulate. Memories are essentially *selective*, *fragmentary* and *imperfect*. If they could *perfectly* simulate previous perceptual experiences they would be *indistinguishable* from them and we would find ourselves *re-living* them – e.g. we would experience a remembered toothache *exactly* as if it were occurring *now*. Memories are often *unreliable* and Hume, arguably, is right to bracket them with products of the *imagination*. Saying "I *seem* to remember but I might be *imagining* ...", is not uncommon. *All* perceptions would appear to require *ownership* i.e. to be the perceptions of a perceiving *something*.[6] In the case of memories, that something must have the *power* of memory – defined by Gibb (2012) as "the ability to encode, store and retrieve thoughts and sensory experiences". The *storage* of anything requires a storage *medium* – i.e. something in which to *inhere*. Hume confesses, however, that identifying "something simple and individual" in which perceptions might "inhere" or "some real connection" uniting their "distinct existences" has proved "too difficult for my

[4] Hume states that perceptions "succeed each other with *inconceivable rapidity*", thus ascribing to them *vanishingly small durations* and making them not just different from commonly recognised sensory/cognitive experiences but, contradictorily, *imperceptible*. His ontological position has been characterised as one of 'perceptual atomism'.

[5] As such, they can *themselves* be the subject of future memories i.e. it *is* possible to remember remembering something.

[6] It appears unintelligible to regard perceptions as perceiving *themselves*. Hume, tautologically, describes *himself* as 'perceiving perceptions' but does not refer specifically to '*perceptions* perceiving perceptions'. This is *implied*, however, by his definition of the self as 'nothing but a bundle of perceptions'.

understanding". He nevertheless expresses the hope that "others perhaps, or myself, upon more mature reflections, may discover some hypothesis that will reconcile these contradictions" [THN A].

Hume says perceptions compose the *mind*, thus equating it with the *self*. How a mere set of perceptions could possess *powers* and be pre-conditioned to interpret 'constant conjunctions' *causally* is unclear.

19. Hume's hope remains unfulfilled because the 'contradictions' to which he refers arise inevitably from his initial premise that the only 'distinct existences' of which we have certain knowledge are atomistically conceived 'perceptions' and his ensuing attempt to construct out of them the entities, including human 'selves', which we commonly distinguish. In particular, he is faced with the problem of explaining what *connects* the perceptions forming an individual self given that they *differ vastly in character* and, being successive, are *non-concurrent*. Identifying "the relation of cause and effect" as a connecting principle just adds to the problem. Hume argues that causation is not something we perceive and can be understood only as a compulsion of the human mind to replicate in its imagination observed and remembered *constant conjunctions* between 'objects'. He defines a cause as: "*an object precedent and contiguous to another, and so united with it in the imagination, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other*" [THN 1.3.24]. A *mind*, for Hume, thus has powers of observation, memory and imagination and is pre-conditioned to interpret some relationships as causative. He would appear to treat 'the mind' as *synonymous* with 'the self', on the basis that "everything that exists is particular: and therefore it must be our several particular perceptions that compose the mind... The mind is not a substance, in which perceptions inhere" [ATHN]. Whether he calls it a mind or a self, the problem for Hume is to explain how a set of perceptions could have the faculties he ascribes to 'it' – including the power to discover, through the application of an induced notion of causation, its own "personal identity".

Hume agrees with Locke that substance is "an uncertain supposition of we know not what".

20. In stating that the mind is not a substance, Hume clearly rejects Descartes' dualist and Berkeley's monist ontologies. His position, at least with regard to *physical* substance, appears similar to Locke's. Locke says we have *ideas* of three sorts of substances – God, finite spirits and bodies [see EHU 2.27.2] – but cannot know for certain whether these correspond to anything existent. The idea of substance, he argues, "we neither have, nor can have, by *sensation* or *reflection*"[7] and can signify nothing but "an uncertain supposition of we know not what ... which we take to be the substratum or support of those ideas we do know" [EHU 1.4.18]. "So that if anyone will examine himself concerning his notion of pure substance in general, he will find he has no other idea of it at all, but only a supposition of he knows not what support of such qualities which are capable of producing simple ideas in us..." [EHU 2.23.2]. Similarly, Hume argues that the idea of substance cannot be derived from "the impressions of sensation or reflection" and that "we have therefore no idea of substance, distinct from that of a collection of particular qualities, nor have we any other meaning when we either talk or reason concerning it" [THN 1.1.6].

Notwithstanding sceptical doubts, we have to take for granted the existence of 'body'.

21. Hume's conviction that the only 'distinct existences' of which we have certain knowledge are our own perceptions leads him to question the notion of external existence. "We may observe that it is universally allowed by philosophers, and is besides pretty obvious of itself, that nothing is ever really present with the mind but its perceptions or impressions and ideas, and that external objects become known to us only by those perceptions they occasion" [THN 1.2.6]. He accepts, nevertheless, the *practical* impossibility of doubting that external objects or 'bodies' exist independently of any perceptual experience we might have of them. "[The sceptic] must assent to the principle concerning the existence of body,

[7] According to Locke, what minds perceive are *ideas*, either of *sensation* (relating to external things or stuff) or *reflection* (relating to the mind's own internal workings). He rejects the possibility of *innate* ideas, arguing that human minds start out as a blank slates and that all their ideas are derived from *experience*. Even if there *were* such things as innate ideas, the fact of their innateness would not, of itself, provide any guarantee of their *truth* i.e. of their correspondence to any sort of reality.

though he cannot pretend, by any arguments of philosophy, to maintain its veracity. Nature has not left this to his choice, and has doubtless esteemed it an affair of too great importance to be trusted to our uncertain reasonings and speculations. We may well ask, *What causes induce us to believe in the existence of body?* but it is in vain to ask, *Whether there be body or not?* That is a point we must take for granted in all our reasonings" [THN 1.4.2].

Hume's inability to reconcile his embodiment with existing as *just* a bundle of perceptions impels him to separate his 'philosophical' from his 'everyday' life – a philosophically defeatist thing to do!

22. In describing his inability to 'catch himself' (see paragraph 16), Hume takes for granted not only that he *has* a body but also that its *activity* is vital to the existence of his perceptions – these ceasing temporarily when it sleeps and permanently when it dies. His natural sense of embodiment, however, is challenged by his 'reasoning' – premised upon perceptions being 'distinct existences' and influenced by his 'scepticism with regard to the senses' [see THN 1.4.2] – that he is *just* a bundle of perceptions. Consequently, he does not explore the possibility that his body might constitute the elusive 'something' which unites his perceptions or in which they could 'inhere' (see paragraph 18). Amongst many things, he is at a loss to explain:

- what causes perceptions to come into and go out of existence;
- how human selves, if mere bundles of perceptions, are able to interact and intercommunicate;
- why, unless they relate to externally existent objects/stuff, the sensory perceptions of different selves should display the commonality and consistency which they do.

"The *intense* view of these manifold contradictions and imperfections in human reason", complains Hume, "has so wrought upon me, and heated my brain, that I am ready to reject all belief and reasoning, and can look upon no opinion even as more probable or likely than another. Where am I, or what? From what causes do I derive my existence, and to what condition shall I return? Whose favour shall I court, and whose anger must I dread? What beings surround me and on whom have I any influence, or who has any influence on me?" He gains respite from his "philosophical melancholy and delirium" only by immersing himself in everyday activities, after which his speculations "appear so cold, and strained, and ridiculous, that I cannot find it in my heart to enter into them any further" [THN 1.4.7].

A *coherent* account of 'selves' and 'minds' must encompass how we *in practice* live our lives.

23. Each of us, it can be assumed, would insist that we comprise an individual *self* distinct from other selves and possess a *mind* distinct from other minds. To declare "I am *not* an individual self and do *not* have a mind of my own" would, indeed, appear contradictory. We may struggle, nevertheless, to define these terms unambiguously. Whilst confident that we have a *body/brain*, moreover, we are liable to confusion about its relationship with our *self/mind* and whether one can exist without the other. Where the results of our 'philosophising' conflict with the background assumptions implicit in how we live our lives, we have to question the coherence of the former at least as much as the validity of the latter – particularly in the case of philosophical stances which even their own adherents are bound to disregard for all practical purposes in their daily lives, an obvious example being a Berkeley-style immaterialist who:

- *qua* 'philosopher', maintains that her perceptual experience of objects/stuff/people arises *solely* from the blending by her of personalised sensory ideas which God implants in her mind – there being nothing existent but 'spirits' and their 'ideas' (see paragraphs 4-7);
- *qua* 'ordinary person', takes for granted the independent existence, *perceived or not*, of things such as cutlery in drawers, the roots of trees, her own body and body parts and, indeed, whatever underlies the surfaces to which her perceptual experience is largely or entirely confined e.g. the *unperceived* flesh, core and pips beneath the perceived skin – or, more accurately, perceived *part* of the skin – of an apple.

Explaining what we mean by our *self/mind* is conceptually more challenging than explaining what we mean by our *body/brain*. Different people are liable to provide different explanations.

24. Explaining at a *basic* level what we mean by our *body* and body parts including our *brain*, appears conceptually straightforward and demonstrable to an extent by *pointing* at them or, in the case of our internal organs, at their presumed location.[8] By contrast, providing even a basic explanation of what we mean by our *self* or *mind* presents a significant conceptual challenge, an obvious limitation being the absence of anything at which we can point. We might, of course, indicate our self by pointing generally at our body and our mind by pointing specifically at our head/brain but this would be to indicate their supposed approximate *location*, not to *equate* one with the other.[9] The extent of the conceptual challenge involved is evidenced by the *divergent* explanations which different people are liable to provide.

- Substance dualists, such as Descartes (see paragraph 2), would explain that our selves/minds (or souls/spirits as they might also call them) are entities which comprise *immaterial* substance, are connected in an unexplained way during our earthly life to our bodies (comprising *material* substance) and, having an independent mode of existence, can survive bodily death and dissolution.
- Immaterialist monists, such as Berkeley (see paragraphs 4-7) would agree with substance dualists that our selves/minds/souls/spirits are *immaterial* beings. Denying the existence of material substance, however, they would claim that our bodies and body parts are just combinations of sensory ideas (not themselves a form of substance) which God feeds into us – gradually varying and finally terminating their supply, it must be assumed, to produce the effects of bodily ageing and death.
- 'Mitigated sceptics' such as Hume (see paragraphs 16-22) would argue that *practically* we have to accept, but *rationally* cannot justify belief in, the existence of our bodies and body parts. Whilst recognising the dependence of our perceptual experience upon the activity of our *bodies/brains*, they would be unable to explain how these relate to our *selves/minds* – which they regard as nothing but bundles of disparate and fleeting perceptions. They would be equally at a loss to explain what might bind such perceptions together or in what they might 'inhere'.

As already indicated, all three approaches lack coherence and, in particular, the ability to account for everyday facts of our sensory/cognitive experience. The same applies to variants such as *phenomenalism* – associated with, amongst others, John Stuart Mill (1806-73) who, defining 'matter' as *'a permanent possibility of sensation*', purports to embrace Berkeleian idealism but without the need for God.[10]

Both *minds* and *bodies/brains* may be said to exist but in different senses.

25. Escaping the conceptual confusion inherent in the above approaches and identifying a coherent alternative requires us to recognise that the *sense* in which things are said to exist can *differ* and that this applies to *minds* and *bodies/brains*. Ryle (1949) argues as follows. "The belief that there is a polar opposition between Mind and Matter is the belief that they are terms of the same logical type... It is perfectly proper to say, in one logical tone of voice, that there exists minds, and to say, in another logical tone of voice, that there exist bodies. But these expressions do not indicate two different species of existence, for 'existence' is not a generic word like 'coloured' or 'sexed'. They indicate two different senses of 'exist', somewhat as 'rising' has different senses in 'the tide is rising', 'hopes are rising', and 'the average age is rising'. A man would be thought to be making a poor joke who said that three things are now rising, namely the tide, hopes and the average age of death. It would be just as good or bad a joke to say that

[8] This is *not* to claim that we could explain the structures and workings of our bodies or brains – about which most of us have at best hazy, if not inaccurate, notions – still less the micro-physical processes taking place within them.

[9] As noted by John Locke (see paragraph E3), we naturally attribute *location* to our self/mind and relate it to that of our body. Were we to identify a location *within* our body, we would probably point to our *brain* but possibly to our *heart* – reflecting a hangover of the belief, traceable back to Ancient Egypt, that it houses our 'soul' and is thus worthy of special treatment after death (e.g. novelist Thomas Hardy's *heart* is buried in Dorset, the ashes of the rest of his body in Westminster Abbey).

[10] As with much superficially beguiling phraseology, the term 'permanent possibility of sensation' turns out, on examination, to be vacuous and question-begging – particularly with regard to the existential status of 'possibilities', what makes different sensory experiences possible or impossible and what causes a mere possibility to become an *actuality*. Berkeley, unlike Mill, does at least recognise the problem – although, having ruled out the existence of 'matter', is forced to invoke a *deus ex machina*.

there exist prime numbers and Wednesdays and public opinions and navies; or that there exist both minds and bodies."

Human bodies/brains exist as physical objects but, also being *living* things, display distinct attributes.

26. Human bodies/brains may be said to exist in much the same way as physical objects in general exist – i.e. as spatio-temporal aggregations of particles in fields of force, their existence being *revealed* by, but not *dependent* upon, their being observed/detected. Along with other living things, however, they also exist as *complex systems* which: come into existence through a process of reproduction; self-regulate their vital functions; discard old and acquire new material in a systematic way which allows them to alter/grow without losing their structural identity; interact with their environment in ways conducive to their own preservation; eventually cease to function and 'die'. The attributes they display, therefore, differ in crucial respects from those of the bulk of observed objects/stuff (such as rocks, rainwater and rivers).

Human minds exist as *cognitive systems realised within human brains*.

27. As already indicated (see paragraph 24), human *minds* are not objects which can be pointed at or located in time/space, suggesting that the sense in which they are said to exist is different from that in which bodies/brains are said to exist. The word 'mind', moreover, can have different senses depending upon whether it is used as a count or a non-count noun. 'What is mind?' (non-count use) and 'what is *a* mind?' or 'what are minds?' (count use), are *different* questions requiring *different* answers. When used as a *non-count* noun (as in, for example, 'the philosophy of mind') the word refers to the human *faculty* which is evidenced in a range of mental *activities* including sensing, perceiving, remembering, conceptualising, reasoning, imagining, desiring, intending, choosing and initiating purposive action/communication. When used as a *count* noun (as in 'my mind', 'your mind', 'their minds') it refers to the cognitive *systems* (of which there are now almost 8 billion worldwide) which exercise this faculty and which, it is evident, are realised within, and impossible without, human *brains*.

Our experiences of selfhood are a product of our minds. The 'point-of-view' nature of our mental experiences entails a *formal* notion of the self which, in a complex form, accounts for our acceptance of responsibility for our actions and our ability to make plans for the future.

28. Where does this leave the human *self*? Many philosophers, as we have seen, *equate* it with the human mind. But is this legitimate? As already argued (see paragraphs 9 & 10), the *variable* and *intermittent* character of our experience of 'self' is suggestive of its being, along with all other mental experiences, the *product* of the workings of our minds. When imagining, reasoning, deciding, acting etc. we are not generally conscious of 'observing' ourselves to be doing so; we just do it. Of course, this does not rule out the possibility of there being parallel *sub*conscious processes linking our ongoing mental and physical activity to some form of self-awareness. There are times, moreover, when we *are* self-aware, especially when interacting with others, deciding what we should/shouldn't do and, indeed, speculating about our own nature, as we are doing right now! Underlying all of our *purposive* mental/physical activity, arguably, is the presupposition of *agency* i.e. that it can be attributed to an entity which thereby bears responsibility for it. Seale (2004) argues that a 'formal' notion of self is entailed by the fact that the experiences involved in such activity occur from a given 'point of view', this not being something seen or otherwise perceived but "a purely formal requirement necessary to render intelligible the character of [the] experiences". He goes on to postulate a notion of self which, whilst still formal, incorporates a high degree of complexity. "The notion of a self that I am postulating is a purely formal notion, but it is more complex. It has to be an entity, such that one and the same entity has consciousness, perception, rationality, the capacity to engage in action, and the capacity to organise perceptions and reasons, so as to perform voluntary actions on the presupposition of freedom. If you have got all of that, you have a self. Now we can account for a whole lot of other features, of which two in particular are central to our notion of the human self. One is responsibility. When I engage in actions I undertake responsibility and thus such questions as desert, blame, reward, justice, praise, and condemnation make a kind of sense that they

would not make otherwise. Second, we are now able to account for the peculiar relations that rational animals have towards time. I can organise time, I can plan for the future, because one and the same self that makes the plans will still exist in the future to execute those plans".

Continuity of *consciousness* appears necessary to the attribution of personal responsibility over time.

29. Searle's reference to 'one and the same self' raises the thorny issue of *personal identity*. If we are to be held responsible not just for our present but also for our past and future actions we must be deemed, in *relevant* respects, to be the *same* person. But what is relevant? The bodily changes to which we are subject over our lives may well alter our self-images but are not obviously relevant to our moral accountability. Clearly relevant would appear to be the content of the cognitive systems comprising our *minds*. It is hard to escape the feeling that same person somehow survives as long as continuity of *consciousness* is preserved. According to John Locke, "*self* is that conscious thinking thing … which is sensible, or conscious of pleasure or pain, capable of happiness or misery, and so is concerned for it*self*, as far as that consciousness extends" [ECHU 2.27.17] and thus "consists not in the identity of substance but ... in the identity of *consciousness*" [ECHU 2.27.19]. For Locke, therefore, continuity of consciousness appears crucial to personal survival. He argues that "if it be possible for the same man to have distinct incommunicable consciousness at different times, it is past doubt that the same man would at different times make different persons" [ECHU 2.27.20]. Locke thus distinguishes between a *man* (i.e. a *human being* considered as a *biological* entity) and a *person* (i.e. a *conscious* entity with an interrelated set of memories). Although not without its problems, Locke's distinction (or something like it) is one that we do appear to make. For example, we might be inclined to describe someone suffering from severe memory loss (perhaps resulting from Alzheimer's disease) as still the same *human being* but not the same *person* that we used to know. A recurrent issue in law is the rationale for putting people on trial for crimes which they committed many years ago but are now incapable of remembering or comprehending due, perhaps, to dementia.

Despite its difficulties, we are bound for practical purposes to assess the nature of individual human selves and the changes to which they may be subject.

30. *Complete* loss of memory (even for sufferers of Alzheimer's disease) is extremely rare. To varying extents we *all* experience memory loss as well as the inability to retrieve stored memories (which are, in any case, fragmentary and often unreliable). Just how much memory of an earlier time in our lives do we have to lose in order to constitute a 'different person' from then and how much retain in order to constitute the 'same person'? There is, moreover, much more to our minds than memories – including thoughts, feelings, beliefs, desires, intentions and motivations. Over the years, the nature and content of these are subject, for better or worse, to *change* – reflected to a varying extent in changing personality and patterns of behaviour. As with memory loss, a crucial question is just how much change has to take place before someone is viewed as a 'different person'? Most change is *gradual* and, as it occurs, *imperceptible* – although 'Road to Damascus' conversions regarding, for example, beliefs and behaviour are not unknown. We might, albeit loosely and figuratively, say that the person we are now differs from the person we were as an infant, a teenager, a young adult, and so on, but would be at a complete loss to identify any clear points of transition. The *lack* of such points requires us, for practical purposes, to invent relatively arbitrary ones e.g. regarding the minimum age (10 in England and Wales) at which we can be held criminally responsible for our actions. Differences in *mental capacity*, have clear implications for the attribution of moral responsibility and this may change significantly over time. Although *continuums* of capacity are involved making the drawing of boundaries tricky, we do in practice hold some people less, or not at all, responsible for their actions due, for example, to their youth, mental impairment or mental illness. It is generally accepted that, to bear moral responsibility, a person must be able to *understand* the nature and likely consequences of their intended actions and make considered *judgements* about them. For all its difficulty and potential arbitrariness, assessing the nature of individual human selves and the changes to which they may be subject over time is unavoidable if we are to identify the most appropriate ways of steering human behaviour in desired directions.

Sadly, there appears to be no realistic prospect of the self surviving brain death.

31. If human minds can be realised only within human brains, there would appear to be no prospect for the survival of the former upon the dissolution of the latter. Only if we adopt a dualist stance and postulate the existence of spirits/souls (or 'ghosts in the machine', as Ryle calls them) capable of independent existence might we believe that our 'selves' could survive bodily death. But such a stance, as we have seen, lacks coherence. Some wilder speculations include the possibility of the cognitive systems comprising our minds/selves being *replicated* within computer systems linked, perhaps, to artificial bodies. Dennett (1991), for example, argues that: "If the self is 'just' the Centre of Narrative Gravity, and if all the phenomena of human consciousness are explicable as 'just' the activities of a virtual machine realised in the astronomically adjustable connections of the human brain, then, in principle, a suitably 'programmed' robot, with a silicon-based computer brain, would be conscious, would have a self. More aptly, there would be a conscious self whose body was the robot and whose brain was the computer." Leaving aside the radical difference in structure, mode of operation and material composition of brains and computers, the copying/replication of 'selves' raises major conceptual issues, not the least of which is the consequence of making *multiple* copies. Which then would be the surviving self? All or none of them?

Roger Jennings
January 2021

References

Argyle, Michael (1988) *The Psychology of Interpersonal Behaviour*. Penguin Books
Berkeley, George (1710) *The Principles of Human Knowledge.*
Berkeley, George (1713) *Three Dialogues between Hylas and Philonous.*
Dennett, Daniel (1991) *Consciousness Explained*. Penguin Books
Descartes, René (1644) *Principles of Philosophy* [translated by John Cottingham]. Cambridge University Press
Gibb, Barry J. (2012) *The Rough Guide to the Brain.* Rough Guides
Hume, David (1739) *A Treatise of Human Nature.*
Kant, Immanuel (1781) *The Critique of Pure Reason*.
Locke, John (1689) *An Essay Concerning Human Understanding.*
Russell, Bertrand (1912) *The Problems of Philosophy.* Oxford University Press
Ryle, Gilbert (1949) *The Concept of Mind.* Hutchinson
Searle, John R. (2004) *Mind: A Brief Introduction.* Oxford University Press
Swaab, Dick (2014) *We Are Our Brains: From the Womb to Alzheimer's*. Penguin Books

Referencing to specific sections of texts

The following examples should make clear the referencing system used:

[ECHU 2.23.19]	*Essay Concerning Human Understanding:* Book 2; Chapter 23; Section 19
[DHP3 H25]	*Dialogues between Hylas and Philonous:* 3rd Dialogue; Hylas' 25th statement
[PHK 7]	*Principles of Human Knowledge:* Paragraph 7
[PP 1.52]	*Principles of Philosophy* : Part 1; Paragraphs 52
[THN 1.4.6]	*Treatise of Human Nature:* Book 1; Part 4; Section 6
[THN A]	*Treatise of Human Nature* : Appendix
[ATHN]	*Abstract of the Treatise of Human Nature* (anonymous but likely written by Hume himself)