MINDS AND BRAINS

(Paper for session at Kingston Philosophy Café on 19 March 2014)

Note: This paper includes material also contained in a revised and extended version of the paper *Stuff and Nonsense: Berkeley and Immaterialism* presented at KPC on 15 May 2013.

Scope of this paper

This paper attempts to place minds and brains within a *coherent* 'model of reality'. In considering dualist, idealist and realist approaches it addresses questions such as the following. What is the relationship between our minds and our brains? Can consciousness be reduced to neurobiological processing within the brain? Can it be equated with exhibited forms of behaviour or functioning? Do our brains work much like computers and, if so, might computers think and feel? Can we reject the reduction of consciousness to physical processes without having to accept 'dualism' (i.e. belief in the separate existence of 'mind' and 'matter')? If we reject the notion of the mind as a 'ghost in the machine', can we retain any meaningful notion of 'self'. What is the relationship between 'ourselves' as conscious beings and 'external reality'? Can we, in any meaningful sense, be said to create it? Is there any other type of reality we might be said to create?

1. FOUR PHILOSOPHERS AND THE PROBLEMS THEY RAISE

Descartes expounds the dual existence of mind and matter.

1.1 A key figure in the 'philosophy of mind' is the French 'rationalist' philosopher René Descartes (1596-1650) who 'reasons' that there are two distinct kinds of substance. "By *substance* we can understand nothing other than a thing which exists in such a way as to depend on no other thing for its existence" [PP 1.51]. "Each substance has one principal property which constitutes its nature and essence and to which all its other properties are referred. Thus extension in length, breadth and depth constitutes the nature of corporeal substance; and thought constitutes the nature of thinking substance" [PP 1.53]. He thus differentiates *extended substance* (comprising infinitely divisible 'stuff', some configured as 'things') from *thinking substance* (comprising indivisible 'spirits' or 'minds'). Apart from 'God', he regards humans as the only spirits, each constituting a separate 'self'. Spirits have *direct* knowledge of themselves through a form of introspection.[1] They have only *indirect* knowledge of the world of 'stuff' by virtue of being joined to the stuff that comprises their 'bodies'. Sensory messages from the outside world travel through their bodies to their brains where they are somehow detected and interpreted by their minds.

Locke argues that 'matter' is a supposed substratum of our sensory 'ideas'.

1.2 The English 'empiricist' philosopher John Locke (1632-1704) also expounds a form of dualism. He argues that once external things have impinged upon our senses, "some motion must be thence continued by our nerves or animal spirits by some parts of our bodies to the brains or the seat of sensation, *there to produce in our minds the particular ideas we have of them*" [EHU 2.8.12]. According to Locke, what minds perceive are *ideas*, either of *reflection* (about the mind's own internal workings) or of *sensation* (about external things or stuff). However, if all that minds perceive through the senses are ideas (such as shapes, colours, textures, tastes, smells and sounds), the existence of 'substance' or 'matter' can be only *inferred*. Locke accepts that the idea of substance can signify only "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]. "The mind being … furnished with a

[1] Famously, Descartes claimed certainty of his own existence through such introspection when he wrote: "Je pense donc je suis", translated into Latin as 'cogito ergo sum' and English as "I think, therefore I am" (perhaps better rendered as "I am thinking, therefore I exist").

great number of the simple ideas conveyed in by the *senses* … takes notice also that a certain number of these simple ideas go constantly together; which being presumed to belong to one thing [and] not imagining how these simple ideas can subsist by themselves, we accustom ourselves to suppose some *substratum* wherein they do subsist and from which they do result; which therefore we call *substance* [EHU 2.23.1]. Locke regards some sensory ideas (solidity, extension, figure, motion, rest and number) as relating to the *primary* qualities of things (i.e. features they *really* possess) and others (e.g. colours, sounds and tastes) to their *secondary* qualities (i.e. powers to produce various sensations in us by virtue of their primary qualities) [EHU 2.8.9-10]. Locke appears confident that material substance, in spite of being an 'uncertain supposition of we know not what', does actually exist and that we know something about its nature by virtue of ideas relating to its primary qualities.

Berkeley says only 'spirits' and their 'ideas' exist. 'Things' are just 'collections of ideas'.

1.3 The Irish philosopher George Berkeley (1685-1753) accepts uncritically Locke's doctrine that the objects of human knowledge are 'ideas'. Those relating to our experienced world of things and stuff he calls 'sensations'.[2] These, he argues, cannot be caused by a hidden substratum of 'matter' as ideas exist only in the mind and cannot be "copies or resemblances" of things existing outside the mind in an "unthinking substance" because "an idea can be like nothing but an idea" [PHK 8]. He defines that which knows or perceives 'ideas' as "mind, spirit, soul, or myself" [PHK 2] and declares it "evident there is no other Substance than Spirit, or *that which perceives*" [PHK 7]. Ideas can be produced only by the 'will of a spirit' but we, as 'finite spirits', cannot determine at will the sensations we experience. Therefore they must be 'imprinted' in our minds by an 'infinite spirit' (God). All things or stuff exist only as 'collections of ideas' in the minds of finite spirits or of God. Although Berkeley is generally lumped together with Locke and Hume as one of the 'British empiricists' who regard *experience* as the fundamental source of human knowledge, his rejection of 'matter' makes him also an *immaterialist*, his belief in a single type of substance a *monist* and his identification of 'ideas' as the sole objects of human knowledge an *idealist*.

Kant claims that we experience a 'phenomenal world' of appearances, determined by our inbuilt 'forms of perception'. Behind this lies a hidden 'noumenal world' of stuff 'as it is in itself'.

1.4 The German philosopher Immanuel Kant (1724-1804) argues that the nature of human minds is such that we can interpret our sensory experience of the world only in terms of subjective 'forms of perception'[3] which are '*a priori*' (i.e. exist *prior* to such experience, being somehow 'hard-wired' into us) not '*a posteriori*' (i.e. formed *after* and *out of* such experience) [Kant, 1781]. Such forms are *empirically real* (i.e. genuine features of our experienced '*phenomenal*' world) but *transcendentally ideal* (i.e. superimposed by our minds upon the '*noumenal*' world of 'things as they are in themselves'). Whereas Locke considers that the perceived primary qualities of things reveal something about their nature, Kant believes that *nothing* in our perceptual experience can be taken to represent anything about the noumenal world. It remains an eternal mystery to us, an even more 'uncertain supposition of we know not what' than Locke's 'substance'.[4] Kant, it is important to

[2] The Scottish philosopher David Hume (1711-76) defines *impressions* as "all our sensations, passions and emotions as they make their first appearance in the soul" and *ideas* as the "faint images of these in thinking and reasoning" [THN 1.1.1]. The generic term he uses to cover both Impressions and ideas is *perceptions*. In his terminology, therefore, we 'perceive our own perceptions'.

[3] These comprise the forms of '*space*' and '*time*'. Kant revered Newton and, like him, maintains that space and time are to be conceived in 'absolute' terms i.e. not purely in terms of the relationship between identifiable 'things' or 'events'. The German philosopher Gottfried Leibnitz (1646-1716) took an opposing view arguing that "space … is something *merely relative*, as time is" and that space is "an *order of coexistences* as time is an *order* of successions".

[4] Berkeley rejects as incoherent the postulation of a world inherently inaccessible to our senses. "For Berkeley the assumption of an unperceived existent is a contradiction in terms. Not so for Kant. What for him is a contradiction in terms is merely the assumption that an existent can be perceived as it is" [Körner, 1955].

emphasise, is not saying that we *choose* the nature and content of our experienced world. *How* we experience things is *determined* by our *a priori* forms of perception. *What* we experience (e.g. whether we see an apple rather than an orange or one apple rather than two) is equally unchosen and presumably must bear *some* relationship (albeit a totally obscure one in Kant's philosophy) to a noumenal world of 'things as they are in themselves'.[5]

There are both similarities and differences in the positions of the four philosophers.

1.5 Descartes, Locke, Berkeley and Kant all share a belief in the existence of minds (*human* minds, at least) which somehow come into existence and then continue indefinitely. All except Berkeley believe that minds perceive, at least 'indirectly', independently existing 'stuff'. Their 'perceptual route' is thus *stuff → senses → brains → minds.* Berkeley, by contrast, believes that nothing exists except minds and their ideas. For him, brains are just passive collections of ideas and cannot, therefore, transmit or generate ideas.[6] His perceptual route is thus *God → sensations → minds.* Kant believes that minds impose their own pre-existing 'forms' upon sensory input and thereby determine the nature of their experienced world. Berkeley says that minds 'blend or combine together' sensations to 'compose objects' but does not explain what guides the blending process. All four philosophers seem to regard minds as unstructured entities (Berkeley describes a 'spirit' as "one simple, undivided, active being") whose internal operations are somehow 'self-illuminating'.

Elements of 'default positions' can be found in both dualism and idealism.

1.6 Searle (1999) argues that on most major philosophical issues there are 'default positions' that we hold 'pre-reflectively' and to depart from which "requires a conscious effort and a convincing argument". They include the following (the first two of which Searle considers true and the third false).

- "There is a real world that exists independently of us, independently of our experiences, our thoughts, our language.
- We have direct perceptual access to that world through our senses, especially touch and vision.
- Each of us consists of two separate entities, a body on the one hand, and a mind or soul on the other … These are joined together during our lifetimes but are independent to the extent that our minds or souls can become detached from our bodies and continue to exist as conscious entities even after our bodies are totally annihilated".

Of our four philosophers, only Berkeley would appear to disagree with the first of these propositions. He believes in a 'real world' but one consisting of God-implanted sensations that exist only when perceived. They all envisage, in different ways, a perceptual role for 'the senses' but one that is less than 'immediate' or 'direct'. For Berkeley 'the senses' relate simply to different types of implanted sensations which minds combine into 'collections'. Locke erects a 'screen of ideas' in front of a 'substratum of matter' and Kant a 'phenomenal world' forever concealing a 'noumenal world' of 'things as they are in themselves' or of 'stuff as it is in itself'. For Kant, all we can ever talk about is the phenomenal world, the structure of which is determined by our in-built 'forms of perception'. All four philosophers accept the existence of minds or souls but struggle to explain how they come into existence, their attachment during 'life' to 'bodies' (which for Berkeley are just 'collections of ideas') and their eventual 'detachment' through 'death'.

---

[5] In Kant's ontology, as pointed out by the German philosopher Arthur Schopenhauer (1788-1860), concepts relating to 'number' (including 'singularity' and 'plurality') apply only to the 'phenomenal' world of our sensory experience. With regard to an intrinsically unknowable 'noumenal' world, therefore, it is unclear whether we should refer to 'stuff as it is in itself' or to 'things as they are in themselves' or, indeed, whether we can say *anything at all*.

[6] "Besides spirits, all that we know or conceive are our own ideas. When therefore you say all ideas are occasioned by impressions in the brain, do you conceive this brain or no? If you do, then you talk of ideas imprinted in an idea, causing the same idea, which is absurd" [DHP2 P11].

Both dualism and idealism run into major conceptual problems.

1.7 A key problem for dualists is to explain the nature of the relationship between two fundamentally different kinds of substance. If mind and matter are different substances how can they possibly interact? Why and how are minds 'shackled' to bodies during 'life' but liberated as 'free spirits' after 'death'? [7] What happens to the mind while the body sleeps a dreamless sleep? How are mental decisions translated into bodily actions? Berkeley might appear to avoid the problems of dualism with his belief in a *single* substance (i.e. mind or spirit) but his idealism gives rise to issues at least as challenging. If 'things' (including our own bodies) exist only as perceived 'collections of sensations', they must be "annihilated and created anew" [PHK 45] whenever we stop or start perceiving them. Berkeley avoids this conclusion only by claiming that God perceives all sensations at all times, thereby ensuring their continuing existence. The sensations implanted by God, it should be noted, must be varied from one mind to another to give each its own unique perspective on a world of 'perceived things'. A crucial problem for Berkeley is how intentions (e.g. to move something) get translated into appropriate sensations in relevant minds, given that sensations can be implanted only by God. If the intentions of different minds *conflict*, moreover, God must choose which to ignore and which to realise in the form of implanted sensations. That different minds want different things is generally ignored by those idealists who claim we somehow 'create our own reality'.

Can we dismiss 'the mind' as just a 'ghost in the machine'?

1.8 Ryle (1949) regards the distinction made by Descartes between mind and matter as a myth involving a 'category mistake' and describes it as 'the dogma of the Ghost in the Machine'. He argues that "the phrase 'in the mind' can and should always be dispensed with. Its use habituates its employers to the view that minds are queer 'places', the occupants of which are special-status phantasms". Most if not all of us, however, would insist that we have minds as well as brains. Can we find a meaningful concept of the mind that avoids the pitfalls of both dualism and idealism?

2. MINDS AND BRAINS EXPLORED AND DEFINED

Flows of consciousness are experienced separately and uniquely by each mind and constitute *real world phenomena* possessing *subjective, first person ontology*.

2.1 There is much scope for conceptual confusion in the word 'mind'. It can operate both as a mass-noun and a count-noun.[8] When used as a *mass*-noun (e.g. when we ask 'what is mind?'), it can suggest a kind of 'substance' or 'stuff'. When used as a *count*-noun (e.g. when we refer to '*a* mind' or 'minds'), it suggests distinct entities or 'things'. Combining the two, we might be tempted to regard individual minds as *bounded* 'mind-stuff'. However, the notion of 'minds' as dimensionless and

[7] Locke feels bound to attribute 'mobility' to 'spirits' since these share the movements of the 'bodies' to which they are tied. "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" [EHU 2.23.19-20].

[8] An 'object' can be defined as 'bounded stuff' (e.g. a brick) and 'substance' as 'sort of stuff' (e.g. clay). The distinction is *reflected* in our use of 'count nouns' and 'mass nouns' but this does not mean that it is linguistically *determined*. Pinker (2007) refers to psychological research which demonstrates the ability of two-year olds ("an age at which children show no signs of distinguishing count nouns and mass nouns in their speech") to differentiate objects from substances. *What* objects or substances we differentiate depends upon our level of focus and the boundaries that we recognise (e.g. we may focus upon a blade, a patch or an entire field of grass). Pinker suggests that "the count-mass distinction … is best thought of as a cognitive lens or attitude by which the mind can construe almost anything as a bounded, countable item or as a boundariless, continuous medium".

positionless 'chunks' of 'mind-stuff' possessing 'non-spatial' boundaries (whatever they might be) is incoherent. An arguably *coherent* mass-noun use of the word 'mind' is simply to refer to 'consciousness' (i.e. the on-going *experience*, familiar to us all, of perceiving, thinking, feeling, etc.). When talking about '*a* mind' or 'minds' (i.e. using the word as a *count*-noun), however, we appear to 'have in mind' not any transient flow of perceptions, thoughts or feelings but *that which experiences, generates or possesses* such a flow. Perception requires a *perceiver*, thought a *thinker* and feelings *something* that feels. As argued by Searle (2004), an intrinsic feature of human consciousness is its *subjective first person ontology*. Consciousness has 'ownership'. My toothaches, for example, are *mine*, not yours or anyone else's. The same applies to my thought processes, memories, emotional states, etc. They exist as a *real world phenomena* pertaining specifically and exclusively to *my* consciousness. Specificity and exclusivity are crucial features of each person's conscious states and the notions of 'mind' and 'self' appear closely related. Talk of 'my mind', 'your mind', 'her/his mind', 'myself', 'yourself', 'her/himself', etc. is both meaningful and commonplace in our daily discourse.

Consciousness appears to be the product of processes occurring within the brain.

2.2 If we reject as incoherent the notion that the possessors of human consciousness are dimensionless and positionless 'chunks' of 'mind-stuff', the obvious alternative candidates are our *brains*[9]. As stated by Searle (2004), "we know for a fact that all of our mental processes are caused by neurobiological processes and we also know that they are going on in the brain and perhaps in the rest of the central nervous system". A molecular biologist's view is provided by Gibb (2012). "Today, most scientists and philosophers agree that, rather than being an essentially separate entity tethered to the brain via a small structure such as the pineal gland,[10] consciousness is an emergent property of the brain as a whole, a natural consequence of millions of neurons processing information in parallel. It may seem astounding that something so 'physical' as electro-biological processes within the brain could produce something so intangible as consciousness, but that is what happens. We just don't yet understand how" (Gibb, 2012).[11] Modern physics countenances, amongst many weird and wonderful things, the equivalence of 'mass' and 'energy', the spontaneous creation and annihilation of 'matter' and 'anti-matter' and the feasibility of the entire universe emerging from, and perhaps returning to, a 'singularity'. It is arguably no big deal to include amongst such wonders the ability of 'stuff', when put together the right way, to generate what we know of, and directly experience as, consciousness,[12] although the somewhat 'mind-blowing' thought occurs that

[9] The cognitive functions of the brain have been only slowly recognised and explored. The ancient Egyptians regarded the heart as the centre of intelligence and emotion, as did the Geek philosopher Aristotle (384-322BC). Thomas Willis (1621-75) was the first to examine the brain with any real scientific rigour.

[10] Descartes believed that the pineal gland, as a single structure located centrally within the brain, provides the 'point of contact' between 'body' and 'soul'.

[11] Gibb describes how the *evolution* of brain structures is evidenced in the human brain, its component parts ranging from the primitive to the most advanced. These include the brain stem or 'reptilian brain', (governing various vital functions), cerebellum (controlling movement and balance), thalamus ('gating', processing and transferring sensory information), hypothalamus (regulating hormone release), basal ganglia (coordinating fine movement), amygdala (generating emotional responses), hippocampus (associated with memory) and the cerebral cortex (described by Gibb as "the crowning achievement of brain evolution, both literally and figuratively" and associated with the 'highest' brain functions including thinking and reasoning).

[12] The commendably undogmatic Locke at one point countenances the possibility that 'matter', if suitably configured, might have the power 'to perceive and think'. "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, superadd to matter a faculty of thinking, than that he should superadd it to another substance with the faculty of thinking …" [EHU 4.3.6]

this entails the ability of 'stuff', in the form of 'brains', to think about the nature of 'stuff' and 'brains' (and of 'thought' and 'consciousness') - as we are now doing.

Consciousness is not *ontologically* reducible to brain processing.

2.3 Evidence of the link between our mental experiences and brain processes (e.g. the parallel impact upon *both* of various legal and illegal drugs) appears incontrovertible and it is difficult to conceive of the relationship other than as *causative*. This is not to say, however, that consciousness can be *equated* with brain functioning. Searle (2004) makes a crucial distinction between *causal* and *ontological* reduction. The fact that conscious experiences may be generated by, and thus *causally* reducible to, neurobiological processes does not mean that they are similarly *ontologically* reducible. A crucial test is whether a description of the one can be substituted, without loss of meaning and reference, for a description of the other. As already argued, an intrinsic feature of human consciousness is its *subjective first person* ontology. My 'first person' description of what I experience when, for example, I eat an apple or watch a football match cannot be substituted by a 'third person' description of the concomitant electro-biological activity within my brain (assuming that such a description were even possible given the billions of neuronal connections and the amazing complexity of the 'parallel-processing' involved). To describe such brain activity is to say *nothing* about my *experience* of eating the apple or watching the match i.e. about what it is actually *like*.

Nor is consciousness ontologically reducible to anything else.

2.4 Although suspicious of 'isms', Searle (2004) names his approach to resolving the 'mind-body' problem '*biological naturalism*'.[13] He emphasises that conscious states are *higher level* features of the brain system (i.e. that "individual neurons are not conscious, but portions of the brain system composed of neurons are") and that conscious states, being "real features of the real world", can *themselves* function causally (e.g. my decision to take a drink causes the arm movement that raises a glass to my lips). We may note that, just as conscious states are ontologically irreducible to brain states, they are similarly irreducible to anything else. 'Behaviourist' attempts, for example, to *equate* conscious experiences (e.g. feelings of pain or pleasure) with outward forms of behaviour (e.g. saying 'ouch' or smiling) are misconceived. 'Third person' descriptions of such behaviour (assuming any is displayed at all) say nothing about the 'first person' experiences involved.[14] Equally misconceived are attempts (e.g. so-called 'Strong AI') to *equate* the operations of the brain with those of a computer (suggesting the brain as analogous to its 'hardware', and the mind to its 'software'). As illustrated by the 'Chinese Room example' devised by Searle (1984), just because a computer might pass the so-called 'Turing Test' by *mimicking* certain operations of the human brain (e.g. outputting correct answers to input questions), this does not mean that what is 'going on' is of the same nature. Crucially it does not mean that computers experience what we experience as *consciousness*. Factually, all the evidence is that 'stuff' in the form of silicon-based digitally-operating machines (unlike stuff in the form of brains) is incapable of generating and experiencing thoughts and feelings or of possessing a notion of 'self'. Indeed, if this is not the case we are in big trouble. The ethical implications would be appalling. If computers *are* conscious beings, then every time we

[13] Searle argues that such a resolution is possible only if we escape the straightjacket of traditional conceptualisations of 'mind' and 'matter' (e.g. Berkeley's conceptualisation of 'matter' as something, if it existed at all, necessarily inert and senseless). Searle states: "I write out of the conviction that the philosophy of mind is the most important subject in contemporary philosophy and that the standard views - dualism, materialism, behaviourism, functionalism, computationalism, eliminativism, epiphenomenalism - are false."

[14] The point is well made by the story of two behaviourists who make love, following which one says to the other "that was great for you, how was it for me?"

scrap a computer we commit a form of murder and humans have been collectively responsible over the years for computer genocide on a massive scale[15].

'Mind' can be used as a count-noun (as in 'my mind' or 'your mind') *without* implying any 'ghost in the machine'. 'A mind' can be defined as a '*cognitive system*' realised within a brain.

2.5 We have suggested that, as a *mass*-noun, 'mind' can be used simply to refer to 'consciousness' i.e. to the product of brain activity. Using the word as a *count*-noun (e.g. when talking about what is 'in our minds'), however, has the potential for conceptual confusion suggesting, as it does, some sort of 'location' or 'container' for our perceptions, thoughts and feelings. This does not mean, however, that we can or should abandon *any* count-noun use of the word. Phrases such as 'in the mind' are, after all, commonplace in our everyday speech and cannot be readily substituted by 'in the brain' or 'in the head' without change or loss of meaning. Arguably we need a word to distinguish the *cognitive system* realised within a brain from both the brain itself and from any particular on-going mental activity. Included within this system are perceptual and intellectual abilities, opinions and beliefs, stored information and memories, personality traits and behavioural tendencies. Defining 'a mind' in this way appears to square with our general count-noun use of the word.[16] Consistent with it are everyday references, for example, to people 'changing their minds', 'bearing something in mind', or even 'losing their minds'. Conceptualising the mind as a *cognitive system* realised within a complex, compartmentalised but interconnected and largely integrated, parallel-processing neurobiological structure such as the brain is consistent with our everyday experience as thinking, feeling, conscious beings. It is consistent, for example, with our experience of different 'levels' of consciousness (including dreaming), awareness of 'self', fluctuating emotional states and evolving personalities. It also helps to explain how 'malfunctions of the system' can occur including lapses of attention, losses of memory and, in extreme cases, the *dis*integration of the system and personality as a result of different parts of the brain failing to connect.

---

[15] If we discover that computers *are* conscious beings it will be necessary, presumably, to draw up a 'Convention' enshrining their 'rights' (including, perhaps, the right to have some say in what tasks they perform for us, the right to decline to be scrapped when we deem they have become obsolete and perhaps even the right to vote). We might have to set up a RSPCCR (Royal Society for the Prevention of Cruelty to Computers and Robots), perhaps appointing a computer or robot as its first President. Ludicrous as all this might sound, these are *serious* consequences if we find out that computers (and perhaps, as suggested by at least one computer/cognitive scientist, even the central heating thermostats in our homes) can, in any meaningful sense, 'think'. The anthropomorphising of a robot in the shape of Marvin, the 'paranoid android', in Douglas Adams' 1978 Radio 4 comedy series *The Hitchhiker's Guide to the Galaxy* has comic effect precisely because it *is* ludicrous. Far from funny, however, is the homicidal computer 'Hal' in Stanley Kubrick's 1968 film *2001: A Space Odyssey*. It is only because we ascribe to 'him' consciousness and an awareness of 'self', that we feel any sense of sympathy (in spite of the fact that he has just 'murdered' four astronauts) as he begs the sole survivor 'Dave' Bowman to stop disconnecting his circuits, claims to feel that his 'mind is going' and to be 'afraid' and finally 'dies' whilst singing Harry Dacre's 1892 song *Daisy Bell (Bicycle Built for Two)*. Searle (1984), it is important to emphasise, is *not* saying that we couldn't, at least in principle, put together something that *does* experience what we know of as consciousness, thought, and feeling - only that the silicon-based digital computer is not 'it'. If we *could* create an intelligent being then all the ethical consequences referred to above become inescapable. The appalling nature of the moral issues involved is graphically illustrated in Mary Shelley's 1818 horror story *Frankenstein; or, The Modern Prometheus*. Related to such issues are those we already face regarding our behaviour towards the members of sentient species other than our own and towards the members of our own species whom we deem to be less than fully capable, due to mental illness or impairment, of 'rational' thought and conduct (the numbers of such people increasing as we live longer and become more prone to memory loss and dementia). For an assortment of 'extracts' relating to 'artificial intelligence' see pages 17-18.

[16] A relevant example of a count-noun use of the word 'mind' is provided by Ron Howard's 2001 film *A Beautiful Mind* starring Russell Crowe and based on the life of mathematician and schizophrenia sufferer John Nash (winner of the 1994 Nobel Memorial Prize for Economics for his work on game theory).

3. MINDS, BRAINS AND 'STUFF'

Our everyday sensory experience of things/stuff is, in any meaningful sense, *direct*.

3.1 The 'scientific model' of perception represents sensory experience as occurring essentially *within* the brain (Gibb, for example, states that "in reality hearing, vision and other sensory impressions occur *inside* the brain") and thus appears to imply that we have 'direct' perceptual access only to such 'internal' experience and can only *infer* the 'external' existence of things or stuff. However, this follows only if we conceive of ourselves as 'spectators' inside our own brains, observing and interpreting a sort of internal 'sensory show'. Dennett (1991) characterises this as the view of the brain as a 'Cartesian Theatre' inhabited by a distinct and separate 'self' and rejects it on the same basis as Ryle (1949) rejects the notion of the self as a 'ghost in the machine'. An alternative is to accept that our sensory experience of things or stuff is just an intrinsic feature of brain activity. What I experience when, for example, I see, pick up and eat an apple is generated by neurobiological processes. There is no 'mini-me' inside my head/brain separately observing these processes or the experiences to which they give rise.[17] The experiences involved are, in any meaningful sense of the word, *direct*. The 'route' by which sensory stimuli reach our brains can be quite complex (including transmission via electronic media) without it being considered 'indirect'. The key requirement for perceptual experience is that there should be some flow of 'sensory information' from observed to observer. How else could things/stuff (in the form of brains) 'observe' other things/stuff if not by generating 'intentional' states about them on the basis of such flows? If my experience of seeing, picking up and eating an apple is not 'direct' and 'immediate' what *more* is needed to make it so? Perceptual experience as an intrinsic feature of brain activity triggered by external stimuli is surely as 'direct' and 'immediate' (if we need to use those words at all) as it gets. It is only if we (falsely) represent our perceptual experience as one in which an independently existing entity called 'the self' observes and interprets an 'internal' sensory show (whether laid on by the brain or by God) that any suggestion of 'indirectness' arises and we end up implying that we somehow 'perceive our own perceptions'. The sensory experience of, for example, seeing an apple just *is* an intrinsic feature of the brain processes involved.

For our everyday perceptual experience, a *realist* stance 'works'.

3.2 At this stage, a not too naïve realist position might be stated as follows.

- We are directly aware through our senses of the existence of 'stuff' (for want of a better word).
- Such stuff exhibits a wide variety of characteristics and forms.
- Some stuff is so structured as to be divisible into hierarchies of distinguishable 'things'.
- Some of these are 'living' things (animal and vegetable) which:
  - maintain their structural identity whilst changing the stuff of which they are composed;
  - respond in some way to stimuli provided by other things/stuff;
  - replicate themselves;
  - eventually cease to function and 'die'.
- Some living things have 'brains' that can process, and trigger reactions to, sensory stimuli.
- In more complex brains such processing may generate 'representations' of things/stuff that help to guide responses to experienced stimuli.
- The most complex brains of which we are aware have cerebral cortexes and the most complex of these are possessed by humans.

[17] This is *not* to say that such phenomena are *ontologically* reducible to such processes (see 2.3). It is important to emphasise that sensory experiences are *real* outcomes of neurobiological processes. Any attempted 'theory of everything' has to explain their occurrence (i.e. the ability of 'stuff', when appropriately configured, to generate everything we know of as perception, thought and feeling).

- Our brains generate 'intentional' states constituting perceptual experience of things and stuff. All stuff, however, exists independently of any such experience and most, including the stuff of which brains are themselves composed, is *never* perceived.
- Our brains produce all that we know of as thought and feeling. Our thought processes enable us to examine the nature of things and stuff and of ourselves as perceiving, thinking and feeling beings. They enable, for example, me to write and you to read this paper.

A test of any philosophical approach is its *internal coherence*.

3.3 It does not take a genius to spot circularity in the above list which starts by referring to the sensory experience of the very beings (ourselves) with which it concludes. It could be argued, however, that such circularity is inherent in *any* attempt to explicate the nature of perception and existence. Inevitably, any such an attempt has to be conducted from within an existing experiential and conceptual framework. We can, nevertheless, test any 'model of reality' for *internal coherence and consistency* (i.e. whether it 'hangs together' on its own terms). The notion of 'stuff' located in time and space does appear both coherent and consistent with the 'world' of our everyday sensory experience. Indeed, it appears *contained* within such experience which is essentially one of things/stuff to which we have perceptual access through our senses but that exist independently of our perceptions (the first two 'default positions' described in 1.6). It is significant that philosophers who profess 'scepticism with regard to the senses' or who, like Berkeley, argue that nothing exists but 'spirits' and their perceived 'ideas', appear to find it impossible to expound their views without using the language of things and stuff. Generally, it can be argued, both sceptical and idealist approaches are *parasitic* upon realist default positions.

Major conceptual issues arise, nevertheless, when we try to *comprehend* the world as represented by our everyday sensory experience.

3.4 Although a broadly realist approach appears to 'work' with regard to our everyday sensory experience, major conceptual issues arise when we seek to comprehend the 'reality' involved. The more we formulate and test alternative 'models of reality' the more we encounter propositions that appear counter-intuitive. Through our senses, we are as much aware of 'forces' as of 'stuff' but the concept of the equivalence of 'mass' and 'energy' is far from intuitive. The nature of space and time, and whether they should be conceived in 'absolute' or 'relative' terms, has long been the subject of philosophic and scientific speculation (see footnote 3). The concept of 'time' (usually treated as a 'fourth dimension' on a par with the dimensions of space but meaningful and measurable only in terms of identifiable *change*) remains especially challenging. At least as conceptually challenging are the divergent postulates (yet to be reconciled) of relativity theory and quantum mechanics. Atoms are no longer regarded as indivisible but the notion of indivisible elementary particles persists. Whilst recognising the problematic nature of the concept of 'infinite divisibility', it still seems legitimate, from the perspective of our 'macro level' experience, to question what 'particles' such as quarks and electrons are actually 'made from' and why they can't be sub-divided, like atoms, into yet smaller particles. Perhaps most counter-intuitive are the postulates of 'string theory' and the notion of 'parallel universes'. Much theorising involves, and appears meaningful only in terms of, mathematical abstraction. However, whilst the postulates of, for example, the standard model of particle physics, quantum mechanics and relativity theory may be fully expressible and understandable only in mathematical terms it does not seem that the 'reality' being described can c*onsist* of mathematical abstractions which exist only as the product of brain activity (and brains seem to be a very recent development in the history of the universe). A crucial problem is how to conceptualise 'entities' which are intrinsically inaccessible to our 'direct' perception. It is tempting to apply concepts and categories meaningful at the level of our everyday sensory experience to hypothesised features lying beyond the direct reach of our senses. To do so, however, may serve only to confuse and mislead.

4. DO WE CREATE OUR PERCEIVED WORLD? NO, WE DON'T.

The 'scientific' approach, with all its conceptual problems, is preferable to simplistic alternatives including the notion that we somehow create what we perceive.

4.1 The conceptually challenging nature of our scientific understanding may heighten the attraction of simplistic alternatives. One such is the notion that we somehow create what we perceive. How we experience the world is, of course, mediated by a wide range of factors including our sensory and cognitive 'equipment', relative position and mental focus. How we categorise and describe things is affected by the conceptual and linguistic framework within which we operate. To accept that our view of reality is 'perspectival', however, does *not* mean that we *create* it, that it has no existence independent of our perceptual experience or that it is determined by our use of language. Searle (1999) puts the point very clearly. "Just as it does not follow from the fact that I see reality always from a point of view and under certain aspects that I never directly perceive reality, so from the fact that I must have a vocabulary in order to state the facts, or a language to identify and describe the facts, it simply does not follow that the facts I am describing or identifying have no independent existence. The fact that there is saltwater in the Atlantic Ocean is a fact that existed long before there was anyone to identify the body of water as the Atlantic Ocean, to identify the stuff in it as water, or to identify one of its chemical components as salt. Of course, in order for us to make all these identifications, we must have a language, but so what? The facts exist utterly independent of language… It is a use-mention fallacy to suppose that the linguistic or conceptual nature of the *identification* of a fact requires that the *fact identified* be itself linguistic in nature. Facts are conditions that make statements true, but they are not identical with their linguistic descriptions. We invent words to state facts and to name things, but it does not follow that we invent the facts or the things."

Philosophers who claim we 'create what we perceive' confuse our ability to *focus* on different things with an ability to *create* them.

4.2 Although patently absurd, the notion of 'creating what we perceive' is espoused by some philosophers. Searle (1999) quotes the American philosopher Nelson Goodman (1906-98) as arguing that "as we make constellations by picking out and putting together certain stars rather than others, so we make stars by drawing certain boundaries rather than others". The comparison between constellations and stars is wholly spurious. We do not regard constellations as anything but trivial patterns picked out from observable stars. We are free to make whatever patterns we like (e.g. to pick out a group of stars to represent 'a plough' or combine them with a larger group to represent 'a bear'). The stars, on the other hand are 'givens'. Only if we could add, remove or shift them (perhaps to make more interesting patterns) could we be said to "make stars". We can't. Goodman's suggestion that where we happen to draw boundaries around stars, or anything else, is simply a matter of choice is equally spurious. A clear constraint is provided by the *content* of our sensory experience. We do not include millions of miles of empty space around a star as *part* of it because that is not what we see. Similarly, the 'boundaries' of the everyday objects that we encounter are *recognised* (rather than 'created') on the basis of what we see and touch. The fact that, *within strict limits*, we can influence the content of our sensory experience (e.g. by altering the focus of our attention) does *not* mean that we thereby create the things or stuff we perceive. Our minds may be able to "construe almost anything as a bounded, countable item or as a boundariless, continuous medium" (see footnote 8) but *how* we so construe anything is tightly constrained by the perceptual content involved. The fact that we can focus upon a blade, a patch or an entire field of grass, for example, does not mean that they are in any way *created* by us. It would represent supreme arrogance on the part of us humans, who comprise a vanishingly small fraction of all the stuff in the universe, to presume that we somehow determine the existence and nature of all other stuff (including, for example, stars that have existed for billions of years before we, or any other sentient beings, emerged upon the scene). If humans are the product of some evolutionary process, and *if*

the existence of anything depends upon it being perceived, we would have to question the existential status of the *unperceived* primordial slime to which we trace back our ancestry, and thus question our own *current* existential status.

Idealism fails to question the nature of what it is that has ideas.

4.3 The proponents of 'idealism' such as Goodman who suggest that we somehow create what we perceive (e.g. the stars, the moon, oceans, mountains, plants, animals and even our own bodies) by virtue of our own ideas seem to take the existence of 'idea-generating minds' simply as 'givens' without questioning their existential status. What do such minds consist of? Can they themselves consist of ideas? Can ideas generate other ideas or be somehow self-generating? Idealists might be asked a revised version of Berkeley's question (see footnote 6). "When you say all things are ideas conceived by minds do you conceive these minds or no? If you do, then you talk of ideas conceived by ideas, causing the same ideas, which is absurd."

Attacks on 'external realism' appear to be motivated by an 'urge to power'.

4.4 Searle (1999) considers the basic claim of 'external realism'[18] (i.e. "that there exists a real world that is totally and absolutely independent of all of our representations, all of our thoughts, feelings, opinions, language, discourse, texts, and so on") as "so obvious and such an essential condition of rationality and even of intelligibility" that it is hard to understand why anybody in their right mind should wish to attack it. It *has*, however, come under attack from modern versions of idealism, "each typically more obscure than the last … under such labels as 'deconstruction', 'ethnomethodology', 'pragmatism' and 'social constructionism'". Searle considers "that as a matter of contemporary and cultural history, the attacks on realism are not driven by arguments, because the arguments are more or less obviously feeble" (how feeble has been illustrated in the last few paragraphs). The deeper reason for the persistent appeal of antirealism is that "it satisfies a basic urge to power. It just seems too disgusting, somehow, that we should have to be at the mercy of the 'real world'. It seems too awful that our representations should have to be answerable to anything but us… In universities, most notably in various humanities disciplines, it is assumed that, if there is no real world, then science is on the same footing as the humanities. They both deal with social constructs, not with independent realities. From this assumption forms of postmodernism, deconstruction, and so on, are easily developed, having been completely turned loose from the tiresome moorings and constraints of having to confront the real world. If the real world is just an invention … then let's get rid of the real world and construct the world we want. That, I think, is the real driving psychological force behind antirealism at the end of the twentieth century."

5. THE NATURE OF CONSCIOUSNESS AND THE 'SELF'

Our perceptual experience is *continuous* and subject to on-going *revision*. The role of *time, memory* and *expectation* is crucial to an understanding of the nature of consciousness.

5.1 Our sensory experience is essentially *continuous* in nature. The multi-processing of sensory inputs by our brains is *dynamic* and subject to *constant revision*. Dennett (1991) encapsulates this in his 'Multiple Drafts' model of consciousness. "All varieties of perception - indeed all varieties of thought or mental activity - are accomplished in the brain by parallel, multi-track processes of interpretation and elaboration of sensory inputs. Information entering the nervous system is under continuous 'editorial revision'… These yield, over the course of time, something rather like a narrative stream or sequence, which can be thought of as subject to continual editing by many processes distributed around the brain". *Time*, for all its conceptual problems, cannot be ignored in

[18] The word *'external'* is ambiguous and invites the question: 'external to *what*?' In the sense used by Searle, 'external reality' covers what exists *independently of being perceived*. It *includes* the stuff of which our *bodies* and *brains* are composed as this exists as independently of perceptual experience as any other stuff.

any understanding of consciousness. [19] Crucial to consciousness is *memory*, without which we would be unable to 'do anything' with fleeting sensory experience. How, for example, could we understand speech or appreciate music if our awareness at any moment was only of immediately experienced syllables, words or notes? Involved in the interpretation of sensory input must be the triggering, via neuronal networks, of memories and emotional responses together with object/substance recognition linked to stored 'patterns' within the brain. *Expectation* plays a key role in determining the content of our consciousness. To interpret our sensory input (relating not only to things/stuff but also to language and music) we have to *anticipate*, to an extent, what follows. This exposes us to the possibility of error (e.g. finding the menacing figure lurking in the mist to be, on closer inspection, just a bush) and demonstrates the need for continuous review and revision.[20] The cognitive sciences have contributed significantly to our understanding of the highly complex processes involved but much remains obscure. This might tempt us to seek simplistic alternative 'explanations' (including variants of idealism) but the temptation must be resisted if we are not to waste time and mental energy on idle metaphysical speculation divorced from reality.

The 'self' is the product of brain activity and self-images generated by the brain at different times (as well as emotional and behavioural responses) may vary significantly.

5.2 It might seem that by exorcising 'the ghost in the machine' (see 1.8) we 'spirit away' the 'self' and that this invalidates what we have said about its meaningfulness (see 2.1). The basic answer to such an objection has already been given (see 2.5). The mind can be defined as a cognitive system realised within a brain and each such 'system' constitutes a separate 'self'. If "all varieties of perception … are accomplished in the brain" (Dennett, 1991), it then follows that consciousness of 'self' is as much the product of brain activity as representations of the external world and this raises the possibility that such activity will generate different 'selves' at different times. On first consideration this might seem objectionable but it is, after all, a reality of our everyday experience. It squares with the *varying* 'images' we have of ourselves and the *varying* ways in which we behave/react depending upon the physical/social settings in which we find ourselves. It squares also with the development of our personalities throughout life and the sometimes catastrophic personality changes that can result from brain injury or disease. The extent to which brain processes are *integrated* and linked to stored *memories* (including memories of previously generated self-images) as well as to built-in *'tendencies and triggers'*, explains the substantial measure of consistency in the self-images we produce and in our emotional and behavioural responses at different times and under different circumstances. The lack of such integration in a minority of brains makes possible the existence of multiple 'selves' and can raise complex legal issues concerning personal accountability.[21] Such issues, of course, arise not only with 'dissociative identity disorder' (previously called 'multiple person disorder') but with other forms of brain 'malfunction'. For most of us, inconsistencies of self-image and behaviour are relatively minor but the ability of humans with no 'certifiable' mental condition to behave 'out of character' (e.g. the loving parents who gassed other people's children in Nazi extermination camps) is frightening. The longer we live, moreover, the more of us are likely to experience the disintegration of personality (associated perhaps with Alzheimer's disease) and eventually find 'ourselves' asking despairingly, like King Lear, "Who is it that can tell me who I am?"[22]

---

[19] The concept of the 'present moment' is highly problematic. If such a moment possesses duration (analogous to a spatial dimension), it can be sub-divided *ad infinitum* until it vanishes into virtual nothingness.

[20] Much humour, we may note, is based upon the confounding of expectations.

[21] Fictional depictions of 'multiple selves' include Robert Louis Stevenson's novel *The Strange Case of Dr Jekyll and Mr Hyde* (1886) and Joseph L. Mankiewicz's film *All About Eve* (1950) starring Bette Davis. Richard Fleischer's film *The Boston Strangler* (1968), starring Tony Curtis, is *loosely* based on the real-life case of Albert DeSalvo (about whose guilt some doubts remain).

[22] We form 'images', of course, not only of ourselves but of other people. In some circumstances it might be argued that others know us better than we know ourselves. In discussing 'self-knowledge', Ryle (1949) points

6. 'BRUTE REALITY' AND 'SOCIAL REALITY'

The fact of consciousness has to be incorporated into any model of 'reality'.

6.1 Acceptance that consciousness is 'an emergent property' of electro-biological processes occurring in the brain (see 2.2) challenges our understanding of the nature of 'reality'.[23] To what we already know (or think we know) about 'particles' and 'forces' we must add their ability, in certain configurations, to produce conscious experience. This is not to say that individual neurons, molecules, atoms or sub-atomic particles are capable of consciousness (any more than genes are capable of being 'selfish'). The experiences of perceiving, thinking, feeling, etc. are essentially *macro-phenomena* of stuff configured as biological entities possessing brains. We cannot rule out the possibility that *non*-biological entities (e.g. computers) might be capable of consciousness *simply on the grounds* that their chemical composition is different (e.g. our bodies and brains are composed mainly of water molecules and, unlike computers, contain no silicon). However, the processing that takes place within computers appears to be fundamentally different in *kind* from that occurring within brains. Dennett (1991) suggests "in principle, a suitably 'programmed' robot, with a silicon-based computer brain, would be conscious, would have a self." He fails to suggest, however, what 'suitable' breakthrough in programming would achieve this or to consider, if an 'evolutionary' process of change is involved, whether computers might *already* possess a primitive level of consciousness (perhaps equivalent to that of a worm). Searle (1984) argues that "the computational properties of the brain are simply not enough to explain its functioning to produce mental states… brains are biological engines; their biology matters. It is not, as several people in artificial intelligence have claimed, just an irrelevant fact about the mind that it happens to be realised in human brains… Of course, some other system might cause mental processes using entirely different chemical or biochemical features from those the brain in fact uses [but] f*or any artefact that we might build which had mental states equivalent to human mental states, the implementation of a computer program would not by itself be sufficient. Rather, the artefact would have to have powers equivalent to the powers of the human brain."*

We occupy a *social and institutional world* that *is* of our own creation.

6.2 Searle (1999) emphasises that the 'brute reality of physical particles in fields of force' is '*observer-independent'* i.e. is not the product of perceptual/cognitive activity on the part of humans or other sentient beings[24]. However, we also 'occupy' a *'social and institutional world'* that *is* determined by the content of human mental states and thereby *'observer-dependent'*. It embodies complex patterns of beliefs, attitudes and behaviour and includes phenomena such as family, marriage, property, money, nation states, governments and a wide range of political, economic,

---

out that "there is 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". By observing our behaviour, others may detect the influence of motives (perhaps selfish ones) to which we blind ourselves. However, if we did have the gift, as described by the Scottish poet Robert Burns (1759-96), "to see ourselves as others see us", we might well encounter many *different* images of ourselves, interpretations of our behaviour and assumptions about our motivations (and might be naturally inclined to accept only the most favourable).

[23] As suggested at the end of 2.2, we can't escape the disturbing thought that *we* and our *brains* are *part* of the very reality (being manifestations of particles and forces as much as anything else) that we are trying to comprehend.

[24] *How we individually experience, conceptualise and categorise that reality*, of course, is 'perspectival' (see 4.1 and 5.1) and at least *partly* 'observer-dependent'. It is also an obvious fact that, as *part* of the world, we regularly make choices, realised through bodily actions (e.g. mowing the lawn), which cause external reality to be different from what it otherwise would be. Conceptually most challenging is the 'two-way' causative process suggested in 2.4. If we accept that conscious states are caused by neurobiological processes in the brain but can also *themselves* function causally, it appears to follow that conscious decisions (e.g. to alter the focus of our attention, utter statements or move our limbs) *both* cause and are caused by changing configurations of the 'particles in fields of force' comprising our brains.

social and other organisations (including Philosophy Cafés!). Although 'observer-dependent' these possess what Searle (1999 and 2010) calls "*epistemic objectivity*". In other words, they exist as identifiable facts about human intentional states amenable to examination and analysis. They are generally realised through the attribution of 'status functions' (e.g. powers, rights, duties and expectations) amongst groups of people and thus embody forms of 'collective intentionality' made possible through human intercommunication. 'Collective intentionality' does not imply unanimity. Conflict between and within groups of people (perhaps relating to their relative position within social/political/economic hierarchies) alters the content of mental states, rendering social and institutional reality inherently fragile. For good or ill, it changes, sometimes gradually and more or less peacefully (e.g. the progressive extension of UK voting rights), sometimes suddenly and violently (e.g. the English, American, French and Russian Revolutions). Searle (2010) points out that "all institutional facts have to bottom out in brute facts". The brute reality of human starvation and violent death, for example, can be related to the institutional reality, both now and in the past, of economic inequality, nationalism, racialism and religious bigotry.

The capacity of human minds for *imagination* has both positive and negative effects.

6.3 Perhaps the most important feature of human minds is their capacity for *imagination*. Positively, this enables us to conceptualise and explore possible present and future states of affairs (including the thoughts and feelings of others) and make appropriate choices. It enables us to examine and theorise about the nature of 'external reality'. It also allows us to create *fictions* (the stuff of story-telling, the dramatic arts, etc.) that may include imagined beings (elves, dragons, talking animals, thinking cars, Santa Claus, etc.) never encountered in reality and regarded, except perhaps by the very young, as mere fantasies. As adults, unfortunately, we remain prone to believe in the independent 'external' existence of some of our own creative fictions (however vague, ill-defined or incoherent these might be). Many people genuinely belief in the existence of 'ghosts' or 'evil spirits'. Most appear to believe that they themselves possess 'souls' which continue to exist after 'bodily death', either 'reincarnated' on earth or 'relocated' to some 'heaven' or 'hell' (i.e. the third 'default position' identified in 1.6). Such a belief is usually associated with a belief in the existence of one or more 'gods' (generally conceived in vaguely human form but also in the form of animals or inanimate objects). The absence of any evidence for their external existence is irrelevant to their causative power. They 'exist' to the extent that they form part of many people's mental 'baggage' and, as such, have immense power to influence behaviour. In parts of the world some people are still being killed as 'witches'[25] and many are killing and dying in the name of their own particular 'god' or interpretation of how that 'god' should be worshipped. A negative aspect of human minds, linked to their capacity for imagination, is their proneness to superstition, irrationality and gullibility. This may reflect, to an extent, the compartmentalised structure of human brains (see footnote 11) and the potential 'battle' between the parts associated with emotion and reasoning. The activity of such parts is interconnected and equally 'real' (being realised in the form of constantly changing configurations of the 'particles in fields of force' comprising brains). The complex patterns of activity involved are crucial to our functioning as active beings in the world making choices in response to external stimuli.[26]

[25] Mary Trembles and Susannah Edwards, hanged in Bideford in 1682, were the last people to be executed in England as 'witches' (under the Witchcraft Act of 1604). The Witchcraft Act of 1735 redefined the offence as one of falsely claiming to possess spiritual powers and moderated its punishment to fines or imprisonment. In 1944 Helen Duncan was goaled for 9 months under the Act which was not repealed until 1951 (when it was replaced by the Fraudulent Mediums Act).

[26] The relative impact on moral choice of reason and emotion has long been the subject of philosophical debate. David Hume, when he wrote "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], appears to ascribe too subservient a role to reason. However, the making of choices (whether 'moral' or otherwise) and their translation into action does appear to require some sort of emotive 'trigger'.

The tendency of human minds to 'externalise' their own mental constructions is the source of confusion, in philosophy as much as anywhere else.

6.4 The tendency of human minds to attribute independent 'external' existence to the 'creatures' of their own imaginations applies to many features of our social, institutional and intellectual world. Although beyond the scope of this paper to explore, the following are worthy of mention here.

- *'Organisational entities'* such as nation states, governments, companies and societies are often treated as 'beings' capable of intentionality. Meaninglessly, for example, people may talk about what their 'country' wants and may even be prepared to die for their 'country, right or wrong'.
- Although use of the word 'we' to denote an identifiable group of people (that necessarily includes the user) is perfectly meaningful, we may nevertheless be confused into postulating the existence of some *unitary being* (perhaps labelled 'the People') distinct from the individuals concerned, "some sort of collective mental entity, some overarching Hegelian World Spirit, some 'we' that floats around mysteriously above us individuals and of which we as individuals are just expressions" [Searle, 1999].
- The tendency to 'externalise' our own mental constructions has spawned whole schools of philosophy. Apart from 'Hegelian idealism' (used by 'fascists' of the far right and far left to justify imposing the 'Will of the People' or the 'Will of the State' upon actual people regardless of their actual wishes), we might mention 'Platonic Forms' i.e. the notion that the 'things' we commonly distinguish (e.g. trees, tables and teaspoons) are ephemeral and imperfect representations of abstract 'Ideal Forms' possessing permanent and independent existence 'outside of space and time' (whatever that might mean).

A final thought

6.5 Our long-term survival as a species depends upon our ability to channel our collective intentionality in a *positive* direction that breaks down barriers between people. Our highly developed brains give us the capacity to do so. The problem is that we appear, at one and the same time, to be both too clever and too stupid for our own good. We are clever enough to understand much about the nature of 'forces' and particles' but stupid enough to use that knowledge to produce, for example, nuclear weapons. It is quite possible that we may prove to be one of the shortest surviving species on the planet (taking most other species with us as we go). As a final thought, we might consider what sort of collective intentionality we could share with intelligent beings living elsewhere in the universe, were we ever to encounter them. One of our four philosophers, John Locke, had the imagination to entertain this possibility. "He that will not set himself proudly at the top of all things but will consider the immensity of this fabric and the great variety that is to be found in this little and inconsiderable part of it which he has to do with, may be apt to think that in other mansions of it there may be other and different intelligent beings of whose faculties he has as little knowledge or apprehension as a worm shut up in one drawer of a cabinet has of the senses or understanding of a man" [EHU 2.2.3].

Roger Jennings
March 2014

References:

Berkeley, George (1710) *The Principles of Human Knowledge.*
Berkeley, George (1713) *Three Dialogues between Hylas and Philonous.*
Dennett, Daniel C. (1991) *Consciousness Explained*. Penguin
Descartes, René (1644) *Principles of Philosophy.*
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.*
Körner, S (1955) *Kant*. Penguin Books
Locke, John (1689) *An Essay concerning Human Understanding.*
Pinker, Steven (2007) *The Stuff of Thought: Language as a Window into Human Nature.* Penguin Books
Ryle, Gilbert (1949) *The Concept of Mind.* Hutchinson
Searle, John R. (1984) *Minds, Brains and Science: The 1984 Reith Lectures.* BBC
Searle, John R. (1999) *Mind, Language and Society: Philosophy in the Real World.* Basic Books
Searle, John R. (2004) *Mind: A Brief Introduction.* Oxford University Press
Searle, John R. (2010) *Making the Social World: The Structure of Human Civilization.* Oxford University Press

Referencing to texts

The following abbreviations are used:

DHP	*Dialogues between Hylas and Philonous*
EHU	*Essay concerning Human Understanding*
PHK	*Principles of Human Knowledge*
PP	*Principles of Philosophy*
THN	*Treatise of Human Nature*

The following examples should make clear the referencing system used:

[DHP2 P11]	2nd Dialogue; Philonous' 11th statement
[EHU 2.7.12]	Book 2; Chapter 7; Section 12
[PHK 2]	Section 2 (of 'Part 1' - the only Part that Berkeley completed)
[PP 1.51]	Part 1; Paragraph 51
[THN 1.1.1]	Book 1; Part 1; Section 1

Robots, Computers and Monsters:
Assorted extracts relating to artificial intelligence
(Supplementary paper for KPC session on 19 March 2014)

Marvin (the manic depressive robot)
Douglas Adams (1979) *The Hitch Hiker's Guide to the Galaxy*

Marvin trudged on down the corridor, still moaning. "… and then of course I've got this terrible pain in all the diodes down my left side …"
"No?" said Arthur grimly as he walked along beside him. "Really?"
"Oh yes," said Marvin. "I mean I've asked for them to be replaced but no one ever listens."
"I can imagine."
Suddenly Marvin stopped, and held up a hand.
"You know what's happened now of course?"
"No, what?" said Arthur, who didn't want to know.
"We've arrived at another of those doors."
There was a sliding door let into the corridor. Marvin eyed it suspiciously.
"Well?" said Ford impatiently. "Do we go through?"
"*Do we go through?*" mimicked Marvin. "Yes. This is the entrance to the bridge. I was told to take you to the bridge. Probably the highest demand that will be made on my intellectual capacities today I shouldn't wonder."
Slowly, with great loathing, he stepped towards the door, like a hunter stalking its prey. Suddenly it slid open.
"*Thank you,*" it said, "*for making a simple door very happy.*"
Deep in Marvin's thorax gears ground.
"Funny," he intoned funereally, "how just when you think life can't possibly get any worse it suddenly does."
He heaved himself through the door and left Ford and Arthur staring at each other and shrugging their shoulders. From inside they heard Marvin's voice again.
"I suppose you'll want to see the aliens now," he said. "Do you want me to sit in a corner and rust, or just fall apart where I'm standing?"
"Yeah, just show them in would you, Marvin?" came another voice.

****

Paranoid android: Cleaning gadget 'switches itself on' and moves onto kitchen hotplate in 'suicide bid'.
Daily Mail 12 November 2013

Firemen were called to a house fire that broke out after a mechanical cleaning gadget somehow switched itself on and destroyed itself by moving onto a kitchen hotplate. Local media in Austria have referred to the incident as 'robot suicide' and even suggested it was fed up with the constant cleaning it had to do. Fireman Helmut Kniewasser was one of those called to tackle the blaze at Hinterstoder in Kirchdorf, Austria. He said: "The home-owner had put the small robot on the work surface to clean up some spilled cereal. Once the robot had done its job it was switched off but left on the kitchen sideboard. The 44-year-old house owner, together with his wife and son, then left the house and were not home when the robot set off. Somehow it seems to have reactivated itself and made its way along the work surface where it pushed a cooking pot out of the way and basically that was the end of it. It pretty quickly started to melt underneath and then stuck to the kitchen hotplate. It then caught fire. By the time we arrived, it was just a pile of ash. The entire building had to be evacuated and there was severe smoke damage particularly in the flat where the robot had been in use. I don't know about the allegations of a robot suicide but the homeowner is insistent that the device was switched off - it's a mystery how it came to be activated and ended up making its way to the hotplate." It took an hour to clean and make the building safe. The family, at least for the moment, is homeless as their apartment is no longer habitable thanks to the smoke damage. The homeowner said: "The company that makes the robots is selling dangerous devices; I intend to sue to get compensation."

****

Robot QT-1 ('Cutie') reasons about 'his' existence.
Isaac Asimov (1941) *Reason* (a short story)

"I have spent these last two days in concentrated introspection," said Cutie, "and the results have been most interesting. I began at the one sure assumption I felt permitted to make. I, myself, exist because I think."
Powell groaned. "Oh, Jupiter, a robot Descartes!"
Cutie continued imperturbably, "And the question that immediately arose was: Just what is the cause of my existence?"
Powell's jaw set lumpily. "You're being foolish. I told you already that we made you."
The robot spread his strong hands in a deprecatory gesture. "I accept nothing on authority. A hypothesis must be backed by reason or else it is worthless – and it goes against all the dictates of logic to suppose that you made me. I say this in no spirit of contempt, but look at you! The material you are made of is soft and flabby, lacking endurance and strength, depending for energy upon the inefficient oxidation of organic material – like that." He pointed a disapproving finger at what remained of Donovan's sandwich. "Periodically you pass into a coma and the least variation in temperature, air pressure, humidity, or radiation intensity impairs your efficiency. You are *makeshift*. I, on the other hand, am a finished product. I absorb electrical energy directly and utilise it with an almost one hundred percent efficiency. I am composed of strong metal, am continually conscious, and can stand extremes of environment easily. These are facts which, with the self evident proposition that no being can create another being superior to itself, smash your silly hypothesis to nothing."

- Powell and Donovan are field-testers of experimental robots made by the 'United States Robots and Mechanical Men Corporation'.
- Cutie concludes that he must have been created by the space station's Energy Converter and begins to worship it as a 'god', calling it 'the Master' and intoning "there is no Master but the Master and QT-1 is his prophet!"
- The world's first robotics company, Unimation Inc. (Connecticut, USA), was founded in 1956 by Joseph

Engelberger who pioneered the manufacture of industrial robots. In 1966, Engelberger and a Unimate robot appeared on Johnny Carson's Tonight Show during which the robot poured a beer, sank a golf putt, and directed the band.

****

2001: A Space Odyssey (released 1968)
Director: Stanley Kubrick
Screenplay: Stanley Kubrick & Arthur C. Clarke

Extract from an interview with members of the spaceship *Discovery* including 'the brain and central nervous system of the ship', a computer called 'Hal':

Interviewer:
"The sixth member of the Discovery crew is the latest result in machine intelligence, the HAL9000 computer, which can reproduce, though some experts still prefer to use the word mimic, most of the activities of the human brain."

Hal:
"The 9000 series is the most reliable computer ever made. No 9000 computer has ever made a mistake or distorted information. We are all, by any practical definition of the words, foolproof and incapable of error… I enjoy working with people … My mission responsibilities range over the entire operations of the ship so I am constantly occupied. I am putting myself to the fullest possible use which is all, I think, that any conscious entity can ever hope to do."

Dave Bowman (scientist aboard Discovery):
"[Hal] acts like he has genuine emotions. Of course, he's programmed that way to make it easier for us to talk to him but as to whether or not he has real feelings is something I don't think anyone can truthfully answer."

****

What a couple of philosophers say.

"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."
Daniel Dennett (1991) *Consciousness Explained*

"The computational properties of the brain are simply not enough to explain its functioning to produce mental states. And indeed, that ought to seem a commonsense scientific conclusion to us anyway because all it does is remind us of the fact that brains are biological engines; their biology matters. It is not, as several people in artificial intelligence have claimed, just an irrelevant fact about the mind that it happens to be realised in human brains… Of course, some other system might cause mental processes using entirely different chemical or biochemical features from those the brain in fact uses. It might turn out that there are beings on other planets, or in other solar systems, that have mental states and use an entirely different biochemistry from ours. Suppose that Martians arrived on earth and we concluded that they had mental states. But suppose that when their heads were opened up, it was discovered that all they had inside was green slime. Well still, the green slime, if it functioned to produce consciousness and all the rest of their mental life, would have to have causal powers equal to those of the human brain… *For any artefact that we might build which had mental states equivalent to human mental states, the implementation of a computer program would not by itself be sufficient. Rather, the artefact would have to have powers equivalent to the powers of the human brain."*
John Searle (1984) *Minds, Brains and Science*

****

Victor Frankenstein constructs a being that has 'mental states equivalent to human mental states'.

"Although I possessed the capacity of bestowing animation, yet to prepare a frame for the reception of it … remained a work of inconceivable difficulty and labour… As the minuteness of the parts formed a great hindrance to my speed, I resolved, contrary to my first intention, to make the being of a gigantic structure; that is to say about eight feet in height and proportionately large… The dissecting room and the slaughterhouse furnished many of my materials… It was on a dreary night of November that I beheld the accomplishment of my toils… I collected the instruments of life around me that I might infuse a spark of being into the lifeless form… I saw the dull yellow eye of the creature open; it breathed hard and a convulsive motion agitated its limbs. How can I delineate the wretch whom with such infinite pains and care I had endeavoured to form? His yellow skin scarcely covered the work of muscles and arteries beneath; his hair was of a lustrous black and flowing, his teeth of a pearly whiteness; but these luxuriances only formed a more horrid contrast with his watery eyes that seemed almost of the same colour as the dun white sockets in which they were set, his shrivelled complexion and straight black lips... Unable to endure the aspect of the being I had created, I rushed out of the room."

Frankenstein's 'monster' curses his creator.

"Hateful day when I received life! Accursed creator! Why did you form a monster so hideous that even you turned from me in disgust? God, in pity, made man beautiful and alluring, after his own image; but my form is a filthy type of yours, more horrid even from the very resemblance… Believe me Frankenstein: I was benevolent; my soul glowed with love and humanity: but am I not alone, miserably alone? You, my creator, abhor me; what hope can I gather from your fellow creatures, who owe me nothing? They spurn and hate me… Shall I not then hate those who abhor me? I will keep no terms with my enemies. I am miserable, and they shall share my wretchedness."

Mary Shelley (1818) *Frankenstein; or, The Modern Prometheus*