

FREE WILL AND RESPONSIBILITY

Prompted by Kingston Philosophy Café discussion on free will held on 9 January 2019.

A few quotes

1. "It's true that there's no such thing as free will. We can't help what we are or what we do. It's not our fault. Nobody's to blame for anything. It's all in your background and ... and your glands. If you're good, that's no achievement of yours – you were just lucky in your glands. If you're rotten, nobody should punish you – you were unlucky, that's all." – Words of architect Pete Keating, a character in *The Fountainhead* (1943), a novel by Ayn Rand (1905-82)
2. "In historical events great men, so-called, are but labels serving to give a name to the event, and like labels they have the least possible connection with the event itself. Every action of theirs, that seems to them an act of their own free will, is in an historical sense not free at all, but in bondage to the whole course of previous history, and predestined from all eternity." – Leo Tolstoy (1828-1910) – *War and Peace* (1869)
3. "Man can do what he wills but he cannot will what he wills." – Arthur Schopenhauer (1788-1860) – *Essays and Aphorisms* (1851)
4. "There once was a man who said, 'Damn! / It is borne in upon me I am / An engine that moves / In predestinate grooves, / I'm not even a bus, I'm a tram.'" – Maurice Evan Hare (1889-1967)
5. "Men at some times are masters of their fates: The fault, dear Brutus, is not in our stars, But in ourselves, that we are underlings." – William Shakespeare (1564-1616) – *Julius Caesar* (1599)
6. "Man is condemned to be free; because once thrown into the world, he is responsible for everything he does. It is up to you to give [life] a meaning." – Jean Paul Sartre (1905-80) – *Being and Nothingness* (1943)
7. "Sir, we know our will is free, and *there's* an end on't." – Samuel Johnson (1709-84) – as recorded in Boswell's *Life of Johnson* Vol. 2 (1769)

To ask if the will is free or unfree is unintelligible. We need to recast the question.

1. One way to dispose of a puzzling question is to deny its intelligibility. Can we do this with the question of 'free will'? As argued, in effect, by philosopher John Locke, to ask whether or not our will is free is to make a *category-mistake*.¹ As generally defined, 'the will' means the *power* to determine our mental and physical actions. It is thus unintelligible to ask if it is free or unfree. A power is *necessarily* something that its possessor is free to use. An 'unfree will' is a contradiction in terms. At issue, therefore, is not whether our will is *free* but whether we *have* a will – in the sense of having the *power* to *choose/originate* our actions. To say that we cannot 'will what we will' (see quote 3) or 'choose what we choose' (or, indeed, 'choose what we choose to choose' – and so on *ad infinitum*) is to claim that we *lack* such a power.

Our experience is one of exercising a *limited* power to make effective choices/decisions.

2. Our everyday experience is one of making *choices* which affect what happens in the world. Most choices (e.g. about how to spend our money or spare time) are of minor significance. Some, especially those made collectively, (e.g. about whether and, if so, how the UK should leave the EU) can have far reaching consequences. Many do not involve bodily action (e.g. choosing the subject of our thoughts or choosing *not* to go for a walk). The obvious response to any claim that we somehow lack the power to make such choices is simply to point to the *experiential fact* that we can and do make them. The power, of course, is *limited* by: the feasibility/difficulty of achieving the intended outcomes; our mental and physical capabilities and predispositions; the presence or absence of precluding, inhibiting or supporting

¹ John Locke (1989) *Essay Concerning Human Understanding*: Book 2; Chapter 21 (*Of Power*):

"The will, in truth, signifies nothing but a power or ability to prefer or choose. And when the will, under the name of a faculty, is considered, as it is, barely as an ability to do something, the absurdity in saying it is free or not free will easily discover itself."

"All the actions that we have any idea of, reducing themselves to ... thinking and motion, so far as a man has power to think, or not to think, to move or not to move, according to the preference or direction of his own mind, so far as a man *free*."

circumstances. The question thus appears to be not whether we *have* a will (i.e. the power to make choices and to act upon them) but what are its *scope* and *limitations*.

If the 'self' is an *illusion*, the question of its freedom of choice does not arise.

3. Those who deny that we have even a *limited* power to make choices which are genuinely our own – including those who argue unintelligibly that this is because we lack 'free will' – do not deny that we *seem* to make such choices but then go on to claim that the experience is an *illusion*. In support of this, they might argue that the power of choice-making cannot attach to human selves because the human 'self' is *also* an illusion. There is no identifiable central controller in our heads directing our thoughts and bodily actions. To postulate such a being is to postulate a '*ghost in the machine*'.² Modern neuroscience appears to lend support to this view. As stated by neuroscientist Michael Gazzaniga: "The view in neuroscience today is that consciousness does not constitute a single, generalised process. It is becoming increasingly clear that consciousness involves a multitude of widely distributed specialised systems and disunited processes the products of which are integrated in a dynamic manner by the interpreter module."³ As its name implies, this module, located in the left hemisphere of the brain, is considered not to *control* the processes but to *interpret* their output in such a way as to weave it into a seemingly coherent whole. The 'self' emerges as a non-continuous but recurrent *construct* of this interpretative system – as evidenced by the *intermittent* and *variable* nature of our experiences of self-awareness.⁴ Notwithstanding its tenuous character, however, consciousness of self appears crucial to our functioning as *agents* who interact with our environment and with each other.

If all conscious experience is the product of brain activity which is either *pre-determined* or *probabilistic*, how can the cognitive systems comprising our minds make choices?

4. Accepting that self-awareness, together with all other conscious experience, is an internally generated feature of cognitive systems (i.e. *minds*) realised within complexly structured brains, still leaves open the question as to whether those *systems* – as opposed to any ghost-like *selves* lurking within them – might be deemed capable of making choices which impact upon the world. The commonly used argument against this possibility is that conscious experience is the *bottom-up* product of brain activity and that brains are composed of stuff which, like *all* stuff, behave either in a *deterministic* or in a *probabilistic/random* way. Psychologist Susan Blackmore, for example, asserts it to be an "obvious fact that everything that happens in this universe is either caused by something that went before or is a truly random event".⁵ If this assertion is a) coherent and b) true, the scope for *any* entity, however composed or constructed, to make its own *system-level* choices appears non-existent.

Reducing all mental activity to brain processing seems to obviate the attribution of *responsibility*.

5. The reduction of everything that we think, say and do to the bottom-up effect of neuronal activity, itself the outcome of the deterministic or probabilistic behaviour of yet lower-level processes, appears to subvert our conventional attribution to individuals of full or partial *responsibility* for their actions. Whilst it may comfort those whose behaviour is *condemned*, it will disappoint those whose behaviour is *commended* to learn that none of us are responsible for anything we do and never, therefore, deserve either blame or praise. Quote 1 provides a simplistic expression of this view⁶ – although its proponents are

² The pejorative phrase "the dogma of the Ghost in the Machine" was coined by philosopher Gilbert Ryle in *The Concept of Mind* (1949) to refer to the myth of "a polar opposition between Mind and Matter" which, he argues, has arisen from the *category-mistake* of treating them as "terms of the same logical type".

³ Michael S. Gazzaniga (2011) *Who's in Charge? - Free Will and the Science of the Brain*.

⁴ In extreme cases the ability of the mind to construct different personas at different moments can become pathological and evidence itself in the form of 'dissociative identity disorder' (previously called 'multiple person disorder').

⁵ See: <https://www.susanblackmore.uk/chapters/living-without-free-will/>

⁶ The words are those of a character in a novel and represent the *antithesis* of the views of its author. Ayn Rand (pen-name of Russian-American writer/philosopher Alisa Zinovyevna Rosenbaum) developed a philosophical system known as 'Objectivism' which rejects altruism, faith and religion in favour of reason and *individualism*. Her celebration of the individual ego as the fount

now more likely to refer to genes or DNA than to 'glands'. Blackmore, confusingly, accepts that when she makes a decision under a given set of circumstances she has to accept both its consequences and the *responsibility* that goes with them but then argues that this arises not because she made the decision of her own free will but because it is "the decision that the whole universe came up with for [her] under those circumstances" – thereby appearing to *shift* responsibility from herself to the *universe*. But the notion of the universe as a conscious being capable of making *decisions* for which it can be held to account, seems absurd. The reduction of human decision-making to the blind workings of lower level processes does not *shift* responsibility – essentially a social construct meaningful only in the context of human interrelationships – it *eliminates* it entirely.⁷ Gazzaniga (2011) portrays a bleak view of the implications of determinism for personal responsibility. "If the universe and everything in it are following predetermined laws, then that seems to imply that individuals are not personally responsible for their actions. Go ahead and eat the Death by Chocolate cake, it was preordained about two billion years ago. Cheat on the test? You have no control over that – go ahead. Not getting along with your husband? Slip him some poison and say the universe made you do it. This is what caused such a stir when Newton presented his universal laws. I call this the Bleak View, but many scientists and determinists think this is the way things are. The rest of us don't believe it. 'The universe made me buy that dress!' or 'the universe made me buy that Boxter!' just isn't going to fly well at the dinner table. If we were to be logical neuroscientists, however, shouldn't it?"

Living as if we have freedom of choice, arguably, is something over which we have *no* choice.

6. Questioning the *logicality* of believing that we are free, within limits, to make choices for which we can be held personally responsible might seem pointless as, arguably, it is something about which we have *no choice* i.e. we are simply *bound* to conduct our lives on this basis. As American philosopher John Searle argues: "Whenever we make up our minds, we have to presuppose freedom. If, for example, I am in a restaurant and I am confronted with a menu and the waiter asks me what I would like, I cannot say 'I'm a determinist, I'll just wait and see what happens', because even that utterance is only intelligible to me as an exercise of my free will... Whenever we decide or act voluntarily, which we do throughout the day, we have to decide or act on the presupposition of our own freedom. Our deciding or acting are unintelligible to us otherwise."⁸ Similarly, it is hard to imagine even the most fanatically determinist of parents teaching their children that, throughout their lives, they will bear *no* responsibility for *anything*, good or bad, which they do and will never, therefore, merit praise or blame. Without the power to make choices for which we take responsibility, there would appear to be no basis for acting *morally*. What's the point of arguing about what we *should* or *shouldn't* do if we believe that what we end up *actually* doing has already been determined or is simply a matter of chance? To engage in such an argument *presupposes* that it might have the effect of *altering* how we would otherwise behave.

How we seek to influence human behaviour assumes the *causal efficacy* of mental experience.

7. In *practice*, we regard moral suasion and the attribution of praise and blame as capable of influencing how people behave. Thus our treatment – by means, for example, of fines, imprisonment, community service and training/education – of people found guilty of criminal behaviour, is intended, at least in part, to *reform* them and also to *deter* them (and potentially others) from such behaviour in the future. Any transformative/deterrent effect, however, could be achieved only if the *mental experiences* associated with their treatment were to cause their future thought patterns – including their beliefs, desires and intentions – to be *different* from what they otherwise would be. The possibility of any such

of all creativity, her promotion of an ethics based upon *self-interest* and her *anti-collectivist* stance has made her a doyen of US social/economic right-wingers – excluding, presumably, those with strong religious beliefs or those discomfited by her negative portrayal in *The Fountainhead* of people who owe their wealth to inheritance rather than personal effort/ability.

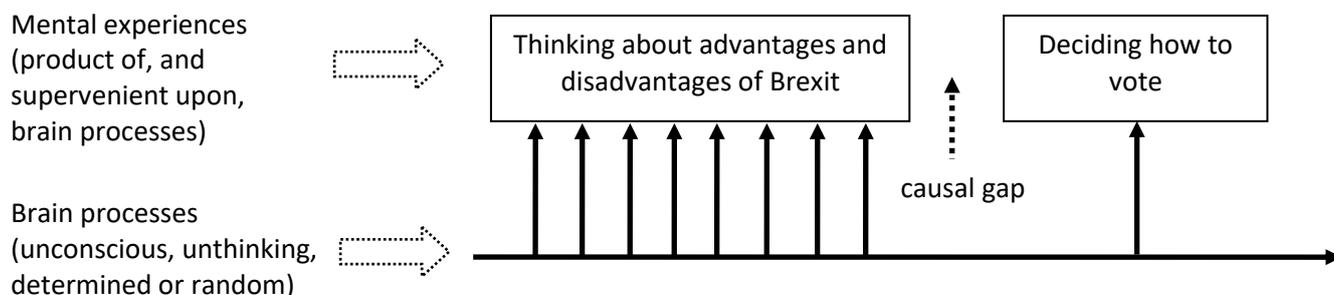
⁷ Philosopher Galen Strawson (see Appendix A) purports to prove the impossibility of our being 'ultimately' responsible for anything we do. The difference between being *ultimately* responsible, rather than just plain *responsible*, for something, however, is obscure. Strawson's pseudo-logical 'proof' appears both specious and irrelevant to the real world of human affairs.

⁸ John R. Searle (2004) *Mind: A Brief Introduction* (Chapter 8 - *Free Will*), Oxford University Press

mental pathway, however, is ruled out if we regard all mental activity as solely the causal result of brain processes, themselves operating deterministically or probabilistically. Mental activity – e.g. thinking about the advantages and disadvantages of the UK's leaving the EU – is then itself causally *inert* and incapable of affecting subsequent mental activity – e.g. deciding, after due consideration, how to vote in the EU Referendum. In philosophical jargon, the term given to this model of the relationship between mental and physical events is *causal supervenience*. Particular mental experiences are regarded not just as invariably *co-existing* with particular brain processes but as the *causal product* of them. Giving a fancy name to such a putative relationship, however, brings us no nearer to understanding how it might work or to resolving the conceptual difficulties involved.

Causal supervenience implies the absence of causal connection at the level of mental experience.

8. To illustrate the nub of the problem, let us consider voters in the 2016 EU referendum. It is safe to assume that all would regard their decision to put a cross in either the 'leave' or 'remain' box of their ballot papers as the causal result of their *thinking*, to an extent at least, about the possible advantages and disadvantages involved. If causal supervenience is a reality, however, *all* their mental experience, including that of thinking about and then deciding how to vote, resulted solely from *unthinking* brain processes (involving neuron firings and lower level particle/wave activity) operating either deterministically or probabilistically. The diagram below provides a simplified representation of the mental experiences and brain processes of a voter.



At the level of mental experience, a causal gap would appear to exist between the exploratory thinking and the ensuing decision. Strictly speaking, although it is difficult to represent in diagrammatic form, there must also be causal gaps between distinct exploratory thoughts – assuming these can be distinguished as *sub-divisions* of *continuums* of mental experience. If they are the *causally inert* product of brain activity, then one thought cannot, at the level of mental experience, trigger or lead to another. Any causal connection between them must operate solely at the *unthinking* level of neuron firings.

A causal gap also appears to exist between a conscious decision to perform a physical act and the actual performance of that act.

9. To complicate matters further, there appears to be an additional causal gap between *consciousness* of making a decision to vote either 'remain' or 'leave' and the *physical act* of casting the vote. If the causal supervenience model is correct, the former cannot *cause* the latter. Each must be the *separate* product of an underlying flow of causally effective, although themselves *unconscious*, brain processes. Of possible relevance here are the experiments conducted by neurologist Benjamin Libet and others since the early 1980s which suggest that the brain activity associated with deciding to make a physical movement starts a fraction of a second *before* conscious awareness of so deciding – although the correct interpretation of the findings of these experiments remains a matter of dispute.⁹ In the case of our voting example, of course, the decision about *how* to vote would normally have been made long before the actual *casting* of the vote (i.e. long before any physical action was involved) and meanwhile would have been stored as an *intention*. The basic issue, however, remains and becomes one of the causal connection between consciously deciding, at some point in time, to put the *intention* into effect and the *physical act* of so doing.

⁹ See: https://en.wikipedia.org/wiki/Neuroscience_of_free_will

The claim of *panpsychism* that *all* entities, from the entire universe down to individual elementary particles, are *conscious* and perhaps even *comprise* consciousness, is *incoherent*.

10. There are philosophers – Searle (2004) calls them *mysterians* – who regard consciousness as intrinsically mysterious, irreducible to brain activity and beyond scientific explanation. Unable to accept that consciousness could emerge from anything *lacking* consciousness, some conclude that *everything*, from the entire universe down to elementary particles such as electrons, must be at least minimally conscious. The name for such a view is *panpsychism*, which translates literally as 'mind is everywhere'.¹⁰ Searle comments that "aside from its inherent implausibility, panpsychism has the additional demerit of being incoherent". Crucially, it fails to explain what determines the *units* of consciousness involved and how they could possibly *interrelate*. How, for example, might the supposed consciousness of a *book* relate to that of each of its *pages* and, ultimately, to that of each of its constituent *molecules, atoms* and *sub-atomic particles*? And what would the conscious experience of any of these be actually *like*? Rather than solving the problem of the relationship between consciousness and brain activity, moreover, panpsychism merely *shifts* the problem to that of the relationship between consciousness and whatever it is that constitutes the virtual infinity of entities claimed to be conscious. A 'hard' version of panpsychism – which might be termed *substance-panpsychism* – avoids this problem by speculating that everything, rather than simply *being* conscious, *comprises* consciousness – but at the expense of lapsing into *total* incoherence. Consciousness is conceived as a *universal substance* of which *everything* is composed. Obvious but unanswered questions which then arise include the following. How can consciousness, conceived as an amorphous substance, fragment into *distinct entities*? What could form the *boundaries* between them? What could an entity *comprising* consciousness be conscious *of*? If all things *comprise* consciousness, how is it possible that at least *some* of them – i.e. *humans* – can lapse into *unconsciousness* e.g. when sleeping a dreamless sleep or if anaesthetised? Conceiving consciousness – essentially an *experiential phenomenon* – as a *substance* (the notion of which raises conceptual issues all of its own) involves a category-mistake resulting inevitably in muddle and confusion.

We know that *some* things are conscious – namely *ourselves*. Consciousness is a necessary but not *sufficient* condition for moral accountability.

11. Whilst the belief that *all* things are conscious is incoherent, the belief that *some* are is perfectly intelligible and, indeed, self-evident. If nothing else, we know *ourselves* to be *perceiving, feeling* and *thinking* beings. The *awareness* associated with such mental activity is a defining feature of consciousness. It varies widely not only in its *nature* and *content* but also in its *intensity* – hence we can talk meaningfully of being *barely, semi, or fully* conscious.¹¹ We are directly aware, of course, only of our *own* mental experience and cannot know for certain what, if anything, is going in *other* people's minds. Having the same basic physiology, however, suggests that the *forms* of consciousness are much the same in all individuals, although precisely how their minds *function* may differ. Differences in *mental capacity*, particularly the capacity to *reason*, have implications for the attribution of *moral responsibility*. 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.

¹⁰ Philosopher Galen Strawson (see footnote 7 and Appendix A) appears to qualify as one of Searle's *mysterians* and also as a *substance-panpsychist* when he states the following. "The nature of physical stuff is mysterious *except insofar as consciousness is itself a form of physical stuff*... What is the fundamental stuff of physical reality, the stuff that is structured in the way physics reveals? The answer, again, is that we don't know – except insofar as this stuff takes the form of conscious experience... There is a fundamental respect in which the ultimate intrinsic nature of the stuff of the universe is unknown to us – except insofar as it is consciousness." See: <https://www.nytimes.com/2016/05/16/opinion/consciousness-isnt-a-mystery-its-matter.html>

¹¹ Deemed culpability for a criminal act may make allow for the perpetrator's conscious state at the time e.g. if it was impaired by extreme tiredness or the effects of medication. An intriguing legal issue is that of 'somnambulist culpability' i.e. responsibility for acts committed whilst asleep! See *Forensic Aspects of Sleep* (1997) by Alexander McCall Smith (with Colin Shapiro).

Non-human mammals can be considered conscious but not *moral* beings.

12. From both their physiology and behaviour, particularly how they interact with their environment, we have good reason to suppose that non-human *mammals* experience some form of consciousness. In the animal world, mammals are unique in possessing a brain with a *cortex*, varying widely between species in size, neuron density and complexity and most highly developed in humans. In humans, the cortex and its interconnections with other parts of the brain – particularly the thalamus – are known to be involved in generating conscious experience.¹² Non-human mammals, to varying degrees, exhibit significant cognitive and problem-solving skills and the power to make basic *choices* – and, in that sense, to exercise a form of *will*. However, whilst some are very close to humans in evolutionary terms¹³, it is assumed that none have sufficient *conceptualising* or *reasoning* ability to qualify as *moral* beings capable of making judgements about what they *should* or *shouldn't* do and of thus bearing responsibility for their actions. In spite of this, we tend to attribute human thoughts and emotions to some of them, particularly to our pets. Animals most likely to be *negatively* anthropomorphised are those deemed a nuisance or danger to humans. Foxes, for example, are characterised by some people as 'wicked' if they kill and eat our chickens – before we get a chance to kill and eat them ourselves!

Evidence points to consciousness as essentially a *neurobiological* phenomenon.

13. We are capable of anthropomorphising not just other *living* things but even *inanimate* objects, including machines such as our cars and computers – particularly when they 'refuse' to do what we want.¹⁴ Some people speculate that computers, in spite of being artefacts rather than life-forms, might experience a degree of consciousness – at least if 'suitably programmed'.¹⁵ Which things might, or might not, be conscious is a *factual* matter to be settled on the basis of *evidence*. All available evidence indicates that consciousness, at least on planet Earth, is a *neurobiological* phenomenon *unique* to a limited range of organisms – most notably humans – possessing complexly structured brains and composed mainly of oxygen, carbon and hydrogen atoms. Digitally-operating silicon-based computers are radically different in structure, mode of operation and material composition. There is *no* evidence of their being conscious (i.e. of *experiencing* thoughts, feelings, pleasure, pain, self-awareness, etc.), which is fortunate as, if such evidence *did* exist, the *ethical* consequences would be horrendous. Complex issues would arise concerning their *treatment*, *rights* and ultimate *killing* (i.e. when we eventually scrap/replace them). Fortunately there is nothing to suggest that our computers/robots are capable, unlike humans, of *any* form of conscious experience e.g. of being happy or sad, wilful or complacent, well-intentioned or malicious, optimistic or pessimistic, etc. All these adjectives are perfectly meaningful when applied to humans but meaningless in relation to machines. Consciousness is essentially an *experiential* and *qualitative* phenomenon.¹⁶ We know what human conscious experience is like. By contrast, there appears to be nothing it could be *like* to be a computer.

¹² Neuroscientist Barry J. Gibb (2012) highlights the importance for consciousness of part of the thalamus known as the *centromedian nucleus* which is wired into a number of different brain regions, including the cortex. It controls levels of arousal and attention and is the target of general anaesthetics. Gibb describes it dramatically as "all that stands between us and nothingness".

¹³ On the evolutionary tree, the branch (hominins) occupied by humans is very close to that occupied by chimpanzees and bonobos, divergence from a common ancestor occurring a mere five or so million years ago.

¹⁴ An extreme, albeit fictional, example is the "good thrashing" that hotel proprietor Basil Fawlty – in the 1970s TV series *Fawlty Towers* – gives his car for 'refusing', once too often, to start.

¹⁵ 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."

¹⁶ As Searle (1999) states: "Conscious states are *qualitative* in the sense that for each conscious state there is a certain way that it feels, there is a certain qualitative character to it. There is something that it is like to drink red wine, and it is quite different from what it is like to listen to music. In that sense, there is nothing it is like to be a house or a tree, because such entities are not conscious."

Issues concerning freedom of choice/action and moral responsibility arise only with *humans*. Whether and, if so, *how far*, an individual is responsible for anything depends upon the meeting of *conditions*.

14. From the above, it would appear that humans *alone* on planet Earth can be held morally accountable for what they do or fail to do – i.e. for their *acts* and *omissions*. For a human agent X to be deemed responsible, at least in part, for event Y, the following conditions need to be met.

- An act or omission on the part of X must have at least *some* effect on Y.
- X must have at least *some* choice over *whether* and if so, *how* to act.
- X must be capable of *identifying* and *evaluating* the likely consequences of pursuing, or failing to pursue, alternative courses of action.

The *extent* of X's responsibility for Y will depend upon how far the above conditions are met, particular account having to be taken of any *constraints* upon X's freedom of choice and of X's capacity to make *informed* and *reasoned* judgements. Where *other* agents also contribute to the occurrence of Y, account has to be taken of the relative significance of X's role, including her/his position in any decision-making or command hierarchy.

Existentialists say we can and should 'transcend' our 'facticity' and be the true authors of our own lives. Doing so, however, might as easily lead to bad as to good results.

15. The philosophical approach known as *existentialism* is noted for the exacting stance it takes towards our responsibility for what we do or fail to do. Existentialists such as Jean-Paul Sartre (see quote 6 and Appendix B) insists that humans have both the ability and the duty to *transcend* (i.e. 'rise above') their *facticity* (i.e. the constraining influence of past experience) and take imaginative/creative control of their lives. They *can* become, in the words of Shakespeare, 'masters of their fates' (see quote 5). They cannot dodge responsibility by transferring it elsewhere. Claims such as "the universe made me do it" or "I had no choice but to obey orders" are just excuses for weaknesses of will. As described by Solomon (1988), "one's facticity is, in particular, one's past, those deeds and events that are over and done with, but whose consequences largely determine our present circumstances, and constitute a significant part of who or what we are." Transcendence, Solomon argues, involves not just *imagining* the various possibilities available to us but *choosing* between them and *acting* upon our choices. "It is the formation of intentions to act that ultimately constitutes our freedom... Facticity and transcendence are the essential components of being human. Facticity defines our situation and who we are – up to that point. Transcendence opens up the world of possibilities – what we can make of that situation and ourselves... Facticity and transcendence are sometimes in brute opposition, as when one tries to change an old habit or way of life, but most of the time they fit together, transcendence 'gearing itself' to facticity, and facticity being reinterpreted according to one's transcendence. Our plans usually fit our circumstances: indeed it is the circumstances that circumscribe if not dictate (but do not determine) our plans." It is important to emphasise that there is no reason why the transcending of facticity should necessarily produce good rather than bad outcomes. For someone conditioned since infancy always to be honest and obey the law, for example, the transcending of facticity might involve taking up a life of crime! Actions have to be judged by their *content* and *consequences*, not their *provenance*. Attributing a particular moral value to something because it happens to have a particular mode of origin exemplifies the so-called *genitive fallacy*.¹⁷

By considering ourselves capable of, and responsible for, choosing/changing our beliefs/behaviours, we adopt, if unwittingly, an existentialist stance. How we behave seems to be affected *positively* by belief in freedom of choice and personal responsibility and *negatively* by belief that everything is predetermined.

16. Whilst only a minority of people would identify themselves as existentialists, there is much about how we *in practice* live our lives that is consistent with existentialism. As already argued (see paragraphs 2 and 6), our everyday experience is one of regularly making choices, most minor but some major, which are

¹⁷ Similarly fallacious is the assumption that acting 'authentically' – i.e. in accordance with one's supposed 'true nature' – is necessarily good. If someone is authentically 'rotten' (see quote 1), might it not be best to encourage her/him to act *inauthentically*?

genuinely *our own* and for which we take *personal responsibility*. Whilst recognising factors which may *constrain* our freedom of choice and push us in certain directions, we do not regard our choices as *dictated* by them – if we did, we would not count them as *choices*. We can accept that our current beliefs, attitudes and patterns of behaviour owe much to the diverse influences to which we have been exposed, particularly during our formative years, without regarding them as fixed and immutable. We can and do modify them from time to time in the light of fresh experience and, crucially, our *interactions with other people*. In all societies, a dialectical process operates whereby people seek to influence each other's beliefs and behaviours – the underlying assumption being that we *can*, within limits, both influence and be influenced by others and are not confined to our own separate 'predestinate grooves' (see quote 4). Whether we do or don't believe this, appears *itself* capable of influencing how we behave. Gazzaniga (2011) refers to psychological studies which "have shown that changing people's sense of responsibility can change their behaviour". In two studies it was found that students encouraged to adopt a deterministic view of life were more likely to cheat in a test, to act aggressively and to be less helpful towards others.¹⁸ It is suggested that "belief in free will may be crucial for motivating people to control their automatic impulses to act selfishly" and that "the mental state supporting the idea of voluntary actions had an effect on the subsequent action decision". Gazzaniga concludes that "it seems that not only do we believe we control our actions, but it is good for everyone to believe it."

If freedom of choice/action and causal determinacy are incompatible, which should we reject?

17. The methods commonly used, whether informally in the *social* sphere or formally in the *legal* sphere, to influence/control beliefs/behaviours are of two broad types – *argument/education* (targeting people's *rational/emotional intelligence*) and *penalty/reward* (targeting their *self-interest*).¹⁹ As previously argued, any transformative/deterrent effect²⁰ attributable to such methods could be realised only through the *mental experiences* they engender – but this route is *negated* if all conscious experience is just a supervenient 'froth' on the surface of predetermined/random brain processes (see paragraphs 7 & 8). We are thus confronted with seemingly *incompatible* presuppositions: a) that we can, within limits, exercise freedom of choice and be responsible for our own actions; and b) that *everything*, including what might *seem* like our own freely chosen actions, are either 'predestined from all eternity' (see quote 2) or mere matters of chance. *Practically* we appear bound to live our lives as if presupposition a) were true (see paragraph 6). On the other hand, the notion of causal determinacy has a strong hold on our imaginations. But is it immune to challenge? Just how obvious, or indeed *intelligible*, is the claim that "everything that happens in this universe is either caused by something that went before or is a truly random event" (see paragraph 4)?

¹⁸ Vohs, K. D. & Schooler, J. W. (2008) The value in believing in free will. Encouraging a belief in determinism increases cheating. *Psychological Science*, 19 (1), 49-54

Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009) Prosocial benefits of feeling free. Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, 35(2), 260-268.

¹⁹ Penalties, whether social or legal, involve imposing on people things they *dislike* and would normally seek to *avoid* (and may thus include the withholding of rewards). Legal penalties generally take the form of fines or loss of freedom. Social penalties generally target people's psychological need for the esteem of others – most often through expressions of disapproval, sometimes through social ostracism.

²⁰ Aside from any reformative/deterrent effect it might have, the deprivation of freedom as a legal penalty serves a *preventative* purpose, rendering criminals incapable of re-offending whilst they remain in confinement. It is generally accepted that the severity of penalties should be proportionate to the seriousness of crimes. Although *means to an end*, however, penalties are not *necessarily* justified by achieving their intended outcomes. A *consequentialist* approach requires that *all* the consequences of actions be taken into account, not just intended ones. On this basis, some penalties might be deemed unacceptable *even if* effective in reducing targeted crimes. A death penalty for murder, for example, might have a significant deterrent effect and is guaranteed to prevent murderers from re-offending but has been rejected by many societies as barbaric. It is certainly a strange way for people to demonstrate their abhorrence at the deliberate taking of a human life by cold-bloodedly doing the same. A major consideration is also the impossibility of restitution if someone is executed but later found to be innocent. This would have been the likely fate of both the 'Birmingham Six' and the 'Guildford Four' – wrongly convicted of the 1974 pub bombings in those cities – had not the death penalty for murder been abolished in Great Britain (1965) and Northern Ireland (1973).

In practice, we treat people as causative agents and use relevant measures to influence/change how they behave – thereby appearing to apply a non-deterministic approach to causation.

18. Some notion of causation appears inescapable in our lives. The key purpose of enquiries into accidents/disasters (such as the Grenfell Tower fire) is to establish their *causes* in order to identify what can be done to prevent their recurrence. Where human acts/omissions are identified as causative factors, relevant measures (e.g. improved regulations, better enforcement and tougher penalties for non-compliance) are routinely introduced with a view to changing how people behave. All of this presupposes that humans have *agency* (i.e. are responsible for making a difference to what happens in the world) and that their behaviour is not predetermined but amenable to influence by equally un-predetermined human interventions. Thus our *practical* approach to the nature of causation with respect to human choices and actions, far from incorporating any notion of causal determinism, appears to *reject* it.

Modern physics appears to espouse causal determinacy and causal indeterminacy.

19. Thomas Pink (2004) defines causal determinism as follows. "Causal determinism is the claim that everything that happens, including our own actions, has already been causally determined to occur. Everything that happens results from earlier causes – causes that determine their effects by ensuring that these effects must occur, leaving no chance for things to happen otherwise. So if causal determinism is true, then at any time what will happen in the future is already entirely fixed and determined by the past." It is widely assumed that causal determinism is a fundamental and unambiguous tenet of the physical sciences. However, whilst many physical processes appear to 'observe', and be predictable in terms of, universal 'laws' of behaviour (more on this later – see paragraph 22), micro-level particle/wave activity is regarded in quantum mechanics as *probabilistic* and incapable of being predicted with exactitude i.e. as *indeterminate*. The stance of modern physics with respect to causation thus appears, ambivalently, to be that *physical behaviour is causally determined except when it isn't* – which begs the question as to what causes some things to exhibit deterministic and others indeterministic behaviour. It is possible, of course, that underlying *seemingly* indeterminate behaviour are determinate processes, our lack of understanding of which means we have to settle for a probabilistic approach to the reality concerned. A more challenging thought is that humans are as much part of the physical world as anything else and thus our brain processes are as much 'infected' at the microphysical level by quantum indeterminacy as the phenomena we observe and seek to comprehend.

Viewing human acts as links in potentially infinite causal chains is problematic.

20. In both the physical and social sciences, major conceptual issues are raised by the notion of causal chains, each link within which has the dual aspect of being the *effect* of its predecessor and the *cause* of its successor. Unclear is the nature of the inter-connection between distinct chains and their temporal extent i.e. whether they stretch infinitely back into the past and infinitely forward into the future or alternatively have points of termination. On the basis of current astrophysical 'wisdom', all of the causal chains to which we are now subject had their point of origin in the mysterious 'singularity' from which the universe supposedly emerged following the so-called Big Bang about 14 billion years ago.²¹ Whether or not they will come to an end at some point in the future depends, it is hypothesised, upon the balance between the forces of expansion and contraction in the universe – an eventual return to a singularity occurring if the latter exceed the former. However, the idea that not just the activity of particles in fields of force but all human choices and actions (e.g. eating the Death by Chocolate cake – see paragraph 5) constitute links in causal chains which, in theory if not in practice, could be traced indefinitely back into the past and projected indefinitely forward into the future, seems preposterous. The point is well made by Helen

²¹ If it is both meaningful and true that "everything that happens in this universe is either caused by something that went before or is a truly random event", why should our backward search for causes stop at a putative 'singularity'? Why should we not ask what caused *it* to come into existence and take the form that it did? Those with religious beliefs who answer 'God' are faced with the same question regarding her/his/its existence and nature. If they confer upon God the status of an 'uncaused cause' they *abandon* the original supposition regarding the *universality* of causality and causal 'chains'.

Steward (2012). "The idea that the evolution of reality over time might depend solely on 'initial conditions' together with purely physical laws, is a really quite extraordinary one... After all, the evolution of reality is profoundly influenced (we tend to think) on a large scale by things such as wars, stock market crashes, global warming, revolutions, industrialization, etc., as well as (on a small scale) by the myriad small decisions each of us makes on a daily basis. To suppose that the occurrence of any of these sorts of things is no more than the high-level manifestation of the inevitable workings-out of the consequences of the initial conditions at the start of the universe (deterministic version) – or else of those initial conditions and merely probabilistic laws, together with nothing more than what may perhaps notionally be thought of as the contribution of mere chance (indeterministic version) – is perhaps one of the most astounding things that has ever managed to obtain the status of philosophical orthodoxy (although it must be conceded that there is strong competition for this title). To believe this would seem to be to consign all sorts of factors that it is natural to regard as causally crucial to the realms of the utterly epiphenomenal. Nothing really matters it would appear, in anything other than an extremely attenuated sense of 'matters', to the unfolding of the world, except the way physical reality was in the beginning, the physical laws, and (perhaps) whatever vagaries are allowed for by the existence of chance. How are we to make room, given this picture, for our basic conviction that we matter to that unfolding, both individually through our actions, and as a species through the phenomena to which our activities have given rise: societies, governments, armies, businesses, religions, technologies, art, literature, science?"

Microphysical equations are *time-symmetric* whereas we conceive causal processes as *time-irreversible*. Viewing causal processes as *chains of events* brings fresh complications.

21. Intrinsic to our everyday notion of causation is *direction* of causality i.e. we view causes as necessarily preceding their effects, their order of occurrence being *irreversible* (e.g. our lives start with our births and end with our deaths, they cannot start with our deaths and end with our births). The mathematical equations used in microphysics, on the other hand, are not similarly time-irreversible. Gazzaniga (2011) quotes systems theorist Howard Pattee as arguing that "the microscopic equations of physics are time-symmetric and therefore conceptually reversible. Consequently the irreversible concept of causation is not formally supportable by microphysical laws, and if used at all it is a purely subjective linguistic interpretation of the laws."²² Gazzaniga acknowledges that "many determinists are anxious to point out that the chain of causes according to determinism is a chain of events not particles, so it never gets down to atoms or subatomic particles. Instead it traces back to the big bang. In Aristotelian terms, the chain is a series of efficient causes rather than material causes". In support of this view, it could be argued that it is the *activity* of particles/waves, not their mere existence, which is causally effective (just as in billiards it is the *striking* of the object ball by the cue ball, not the cue ball *per se*, which causes the object ball to move) and that such activity is best conceived as *chains of events*. It has been suggested, indeed, that what we commonly call *substance* comprises nothing more than grouped events. In discussing relativity theory, Bertrand Russell argues that "relativity demands the abandonment of the old conception of 'matter', which is infected by the metaphysics associated with 'substance', and represents a point of view not really necessary in dealing with phenomena". He suggests that "all the facts and laws of physics can be interpreted without assuming that 'matter' is anything more than groups of events."²³ However, what might constitute a single event and what might bind a number of events together to form a space-time 'group', is far from clear. As commonly conceived, events – whether at the micro or macro level (e.g. whether affecting sub-atomic particles, stars or human beings – are time-extended *happenings* involving activity/change on the part of one or more substantial '*somethings*'. They are not viewed as *constituting* those somethings. We are clearly in murky conceptual waters here, but awareness of the complexities involved should at least caution us against naively assuming, when discussing 'free will', that causal determinism is an obvious and unambiguous tenet of the physical sciences.

²² Pattee, H.H. (2001) *Causation, control and the evolution of complexity*. In P. B. Anderson, P. V. Christiansen, C. Emmeche & M. O. Finnerman (Eds.) *Downward causation: Minds, bodies and matter*, Aarhus University Press

²³ Russell, Bertrand (1925) *ABC of Relativity*. Routledge Classics (Reprint of 1985 4th Edition)

Calling scientific generalisations *laws* gives the false impression that natural phenomena are somehow required to behave as they do. Physics does not treat causation as a distinct factor.

22. In the physical sciences, the mathematical models built upon *observed regularities* in the behaviour of natural phenomena, are commonly referred to as *laws*. Use of this term – borrowed from world of human affairs – is unfortunate as it suggests, misleadingly, that such behaviour is a matter of *requirement* rather than simply an observed fact. Scientist and science fiction writer Isaac Asimov, when explaining Newton's three 'laws' of motion, cautions his readers as follows. "The important generalisations of science are brief descriptions of the behaviour of the universe that are known to cover all observed cases. It is strongly felt that they will also cover all unobserved cases, here or anywhere, now or at any time. Such generalisations are sometimes called 'laws of nature'. This is actually a poor phrase because it seems to draw an analogy with man-made law, as something that is imposed and can be repealed, that can be violated at the cost of a penalty, and so on. All such analogies are misleading. It would be better therefore to speak of 'Newton's generalisations concerning motion'."²⁴ We should also be cautioned against assuming that the key concepts of 'mass' and 'force' – which are fundamental to modern physics and feature extensively in its 'generalisations' – are conceptually straightforward and uncontroversial. Their uncertain nature is reflected in the *interdependent* way in which they have to be defined – force being *whatever it is* that causes a body possessing mass to accelerate and mass being *whatever it is* in a body that causes it to resist the effect of a force. The uncertain nature of the related concept of 'energy' was recognised by American physicist Richard Feynman when he said: "It is important to realize that, in physics today, we have no knowledge of what energy *is*". Although forces are commonly viewed as having causative effects – e.g. tides are deemed to be caused by the gravitational pull of the moon – the mathematical models of physics do not regard causation as a distinct factor e.g. as a fifth force to be added to the four (strong, electromagnetic, weak and gravitational) currently recognised. Any concept of causation – the nature and meaning of which remain problematic – thus appears to serve no obvious purpose in physics, the practical concern of which is simply to model, on the basis of observed regularities, how things behave with a view to predicting how they will behave in the future and in set circumstances. Russell (1925) makes the point as follows. "The physicist, who knows nothing of matter except certain laws of its movements, nevertheless knows enough to be able to manipulate it. After working through whole strings of equations, in which the symbols stand for things whose intrinsic nature can never be known to us, the physicist arrives at last at a result which can be interpreted in terms of our own perceptions, and utilised to bring about desired effects in our own lives. What we know about matter, abstract and schematic as it is, is enough, in principle, to tell us the rules²⁵ according to which it produces perceptions and feelings in ourselves; and it is upon these rules that *practical* uses of physics depend."

Does Kant's phenomenal/noumenal distinction suggest a way in which freedom of human choice/action might be compatible with causal determinism?

23. Russell's reference to the unknowability of the intrinsic nature of things brings to mind the distinction made by Immanuel Kant²⁶ between *phenomena* (i.e. things as they *appear* to us) and *noumena* (i.e. things as they are 'in themselves'). Our *own* nature, argues Kant, is such that we are compelled to *intuit* our raw sensory experience in terms of *a priori*²⁷ 'forms' of *time* and *space* and to *conceptualise* the representations thus formed in terms of *a priori* 'categories' which include *causality-and-dependence*. If

²⁴ Asimov, Isaac (1966) *Understanding Physics: Volume 1 - Motion, Sound and Heat*, Mentor

²⁵ The same objections apply to Russell's use of the word *rules* as to that of *laws* i.e. both imply that things are, in some unexplained way, *required/compelled* to behave as they do.

²⁶ See *The Critique of Pure Reason* (1781) and also *The Critique of Practical Reason* (1788).

²⁷ *Knowledge is a priori* if not derived from, and thus immune to falsification by, empirical evidence (making it, in that sense, *necessary*) and *a posteriori* if derived from, and thus potentially falsifiable by, such evidence (making it *contingent*). Statements are *analytic* if true by virtue of the meaning of their constituent terms and *synthetic* if true by reference to empirical facts. Kant argues, contentiously, for the possibility of 'synthetic *a priori*' judgements i.e. ones which *refer* to empirical matters but which are *known* to be true *independently* of empirical evidence. In addition to mathematical propositions such as $5+7=12$, Körner (1955) cites assertions such as 'Every change has a cause' as examples of what Kant would deem synthetic *a priori* judgements.

causality is an attribute of phenomena but not, or at least not *necessarily*, of noumena – which are, as defined by Kant, intrinsically unknowable – then the apparent incompatibility of freedom of choice/action and causal determinism might be resolved if, 'noumenally', causality either does not exist or is not causality 'as we know it'. Körner (1955) suggests this possibility as follows. "The third antimony²⁸ concerns the question whether there is or is not freedom, i.e. are there or are there not uncaused causes? It is resolved by showing that the thesis – that all phenomena are subject to 'causality' according to laws of nature' – is compatible with the antithesis that a different kind of causality, allowing for uncaused causes, exists for *noumena* or things in themselves. The latter kind of causality is, of course, only an Idea²⁹ – the Idea of freedom – which according to Kant is necessary to account for the experience of moral obligation." A similar suggestion is made by Jane O'Grady (2018). "Kant argues that we do not garner a notion of causal necessity from experience, but that experience is already primed and imbued with the category of 'causation-and-dependence'. Since it is one of the categories we impose³⁰ on the world, it is therefore not an intrinsic part of reality, the true nature of which is unknowable. It is therefore possible that 'I', as well as constituting a 'formal unity' which gives coherence to my experience, may be part of the noumena, to which causality does not apply."

The seeming incompatibility of freedom of choice/action and causal determinism is *not* resolved if causality applies as much to our noumenal as to our phenomenal selves.

24. Kant's division of reality into separate, but somehow related, phenomenal and noumenal realms might appear consistent with our commonplace view of things as existing *independently* of the awareness, *if any*, which humans and other sentient beings may have of them, such awareness being partial, superficial (in the literal sense of being confined mainly to *surfaces*), fleeting and conditioned by the sensory capabilities, cognitive processes and intentional focuses of the beings concerned. These *differ* not only between species but between the members of the same species. By its very nature, the *specific* sensory/cognitive experience of each individual is *unique*. Arguably, therefore, Kant's model of reality should accommodate not just *one* phenomenal realm but as many phenomenal realms as there are sentient individuals – each, potentially at least, imposing *different* sets of forms/categories upon their sensory input. A particular weakness of Kant's formulation is its lack of explanation as to what it is about humans that causes them, individually or collectively, to impose, if they do, the *same* such set. Another weakness is its representation of the noumenal realm as *necessarily* a closed book, in spite of its being the assumed and only source of the sensory input cognitively processed by sentient beings – their resulting perceptual experience thereby possessing sufficient *commonality* of content to cause them to relate it to the *same* sets of things. It is possible, therefore, that the ways in which we perceive/conceptualise such things are not mere subjective 'impositions' but shaped by, and revelatory of, *something* about their real nature.³¹ As Körner (1955) argues: "One can agree with Kant's view that the matter and form of perception are distinct, without sharing his view that the form is subjective. Thus even a realist, who believes that the thing he perceives exists just as he perceives it, could adopt the Kantian distinction without inconsistency." It is at least possible, therefore, that causality-and-dependence is as much an intrinsic feature of noumena as of phenomena and applies as much to our noumenal as to our phenomenal 'selves'. The hope that Kant's phenomenal/noumenal distinction could explain how human freedom of choice/action might be compatible with causal determinism is thus, to say the least, flimsy. The existential nature of human

²⁸ Kant identifies four 'antimonies' (pairs of seemingly mutually exclusive concepts), the third concerning causality and freedom.

²⁹ Körner (1955) explains that Kant uses the term 'Idea' to designate notions which, although "clearly not *a posteriori* (abstracted from perception)... are yet *not the same sort of a priori* concept as, say, causality is. They differ in the important respect that while they are like causality in not being derived from perception, they are yet not, as causality is, applicable to perception."

³⁰ Talk of *imposing* our own forms/categories upon the world might appeal superficially to adherents of the "we create our own reality" tendency in philosophy. In Kant's ontology, however, we do not originate our own sensory experience and such imposing does not affect how things are 'in themselves', only how they are *represented* in human consciousness.

³¹ This is a *scientific*, as well as conventional, view. As Russell (1925) states, "Physicists, like ordinary people, believe that their perceptions give them knowledge about what is really occurring in the physical world and not only about their private experiences. Professionally, they regard the physical world as 'real', not merely as something which human beings dream."

'selves' is, in any case, ambiguous as is the reason why, if *noumenally* unconstrained by causality, such selves, when structuring their *phenomenal* world(s), should impose the category of causality upon the things, including *themselves*, which feature within it.

Kant's noumenal realm provides a convenient repository for *anything* we might care to imagine.

25. An obvious question for Kant is why, if humans are aware of *nothing* but phenomena, any of them should conceive the existence of something *else* – noumena – of which they neither have, nor ever *could* have, any knowledge. Presumably, there must be something in our phenomenal experience which causes us to find a distinction between phenomena and noumena both necessary and intelligible. As argued in the previous paragraph, the content of our everyday experience is such that we are bound to distinguish between our perceptual *experiences* and the *objects* of those experiences – between, for example, our various *sightings* of the moon and the moon *itself*. We readily acknowledge, moreover, the *partial* nature of our sensory experience and countenance the existence of features beyond the direct reach of our senses – most obviously in our speculations about reality at the microphysical level. Such speculations, however, have to be tested against observed phenomena (see Russell's argument in paragraph 22). Kant, on the other hand, so defines the noumenal realm as to place it beyond any such testing. The *physical* phenomena we experience through our senses, he argues, are shaped entirely by our pre-conditioned forms/categories and can tell us *nothing* about any hidden noumenal reality to which they might relate.³² Similarly we cannot know how, if at all, any *non-physical* phenomena identified by us might relate to something noumenal. Kant, nevertheless, is prepared to "deny knowledge in order to make room for *faith*" e.g. to accept that, although lacking *perceptual* evidence for their existence (see footnote 29), we can have faith in the noumenal reality of, for example, 'God', 'souls' and 'free-will'. This is, of course, to make room as much for *superstition* as for faith which, in any case, appears unavoidably *blind*. The noumenal realm provides a convenient 'black box' repository for *anything* we might care to imagine, placing it conveniently beyond the need for, and indeed possibility of, evidence-based argument and justification.³³ Kant's attribution of noumenal reality to a variety of moral precepts – or 'categorical imperatives' as he calls them – represents them as *objective* and thus immune to disagreement, putting them on a par with the imagined commandments of one or more putative 'gods'. Kant, we might note, regards as categorically imperative the prescription that convicted murderers should be put to death – something which is far from universally accepted as an objective and incontrovertible truth (see footnote 20). Kant makes moral dispute a matter of rival and essentially irresolvable claims to privileged insight into the nature/content of a wholly mysterious and eternally inaccessible noumenal realm rather than a dialectical process in which people freely and openly deploy intellectual/emotional arguments relating to the likely consequences, in the one reality of which we are aware, of alternative courses of action.

³² The nature of the relationship is unclear as *none* of the forms/categories applying to phenomena can be presumed to apply to noumena. If there is a noumenal counterpart to our phenomenal moon, for example, it is *meaningless* to ask about its size, location and history. We do not even know whether to refer to noumenal reality as comprising a multiplicity of *things* (noumena) or a single *thing* (noumenon), as the categories of singularity and plurality apply only to phenomena.

³³ In *The World as Will and Representation* (1818), Arthur Schopenhauer (see quote 3) imagines noumenal reality as *comprising* 'will' conceived as an ineluctable, irrational and essentially *amoral* force lacking in purpose other than its own perpetuation and pervading *all* 'things', whether animate or inanimate. He can be included amongst Searle's 'mysterians' (see paragraph 10), his approach closely resembling that of panpsychism – 'will' taking the place of 'consciousness' – and sharing with it all the problems which make both incoherent. Christopher Ryan explains Schopenhauer's line of reasoning as follows. "To Schopenhauer [the] experience of willing is not a representation of the world but *a direct experience of the inner nature of the world*, which he calls 'will'..... If we remove all inessential aspects of the concept of will, so that we are left only with its innermost essence, we might use the word 'will' to mean not only the metaphysical inner reality of humans and animal bodies, but also that of the organic life of plants and the vital strivings of natural forces in nature, such as gravity. By doing so we have gained metaphysical knowledge of the inner reality of the world that lights up in our minds as a spatio-temporal world of causally interacting objects. And this indeed is the path of reasoning that Schopenhauer does take, to establish that the inner reality of the whole world of appearances is itself will." Christopher Ryan (2019) *Poodle as Representation, Rottweiler as Will*, *Philosophy Now*, Issue 134 (pages 6-10), October/November 2019

Kant neglects the power of humans to affect how things *are*, not just *appear*. Modern physics does not accept his *a priori* concepts as indubitable truths.

26. From the above, it can be seen that there are many problems with Kant's formulation – to which the following should be added.

- Humans are cast largely in the role of *observers* who are *compelled* (how is unclear) to interpret their sensory experience in set ways. Their role as *agents* who form part of their environment, interact with it and, within limits, can *change* it – i.e. change not only how things appear but how they *are* – is neglected.
- A basically *static* view appears to be taken of the nature of things, as opposed to a *dynamic* one which recognises that everything is subject to *change* and *evolution* – e.g. the creation of heavier elements such as gold at an intermediate stage in the history of the universe and, crucially, the emergence at a comparatively recent stage of *sentient/conscious* beings capable of intentional choice/action.
- Kant's view of the ways in which humans are compelled to interpret their sensory experience reflects the scientific 'wisdom' of his day. He follows Newton, for example, in regarding space and time as *absolute* rather than relative. Körner (1955) points out that this was challenged during Newton's own lifetime by German philosopher Gottfried Leibnitz who held that "space ... is something *merely relative*, as time is", that space is "an *order of coexistences* as time is an *order of successions*" and that to regard time as "a substance, or at least an absolute being" is "a fancy."
- Modern relativity theory and concepts such as anti-matter, quantum indeterminacy and quantum entanglement call into question the validity of Kant's 'synthetic *a priori*' forms/categories, as well as challenging our conventional notions of the nature of reality/causation. Physicist Werner Heisenberg (1962) suggests that Kant's "central concept of the 'synthetic judgements *a priori*' has been completely annihilated by the discoveries of our century. The theory of relativity has changed our views on space and time, it has in fact revealed entirely new features of space and time, of which nothing is seen in Kant's *a priori* forms of pure intuition... The *a priori* concepts which Kant considered an undisputable truth are no longer contained in the scientific system of modern physics."

Kant's phenomenal/noumenal distinction encourages a 'two-world' view of reality. A 'one-world' approach, however, is needed.

27. Whilst not 'dualist' in the sense of counter-opposing 'mind' and 'matter', Kant's distinction between phenomena and noumena encourages a two-world view of reality. It leaves open and unexplored the relationship between the types of phenomena we commonly distinguish including: physical things/stuff (e.g. stones and sea-water); physical processes (e.g. earthquakes, fires and the ageing of living things); perceptions and sensations (e.g. sights, sounds, smells and stomach-aches); emotions (e.g. feelings of love and loathing); purposive mental states (e.g. desires, aspirations and intentions); memories and imaginings (crucial to rendering coherent our on-going perceptual experience and identifying/exploring future possibilities as well as the stuff of dreams and story-telling); thoughts, arguments and reasonings (such as those contained in this paper); social/institutional entities, practices and activities (e.g. governments, companies, money, marriage, laws and wars). Broadly-speaking, the phenomena involved tend to be categorised as either *physical* or *mental* although there is clearly a close connection between the two e.g. the mental states which lead to wars have very physical consequences. The challenge for both the physical and social sciences (or, more simply, for ourselves as intelligent beings) is to connect and reconcile the two *without reducing one to the other* e.g. without reducing our thought processes to the blind product of pre-determined or random neuron firings (see paragraph 8) or phenomena such as stars, trees, seas and, indeed, our own bodies and body parts to 'combinations of ideas'.³⁴ What appears to be required is a re-conceptualisation of our traditional notions of the mental and the physical. Searle (2004) argues as follows. "The worst mistake is to suppose that the common-sense distinction between mental states naively

³⁴ The classic example of such an approach is that of philosopher George Berkeley who, in his *Principles of Human Knowledge* (1710), conceives a reality comprising nothing but 'spirits' and their 'ideas'. He explains our inability to *will* the sensory ideas we experience (e.g. what we see when we open our eyes) by attributing their source to an 'infinite spirit', namely 'God'.

construed and physical states naively construed is an expression of some deep metaphysical distinction... The problem is that the terms have traditionally been defined so as to be mutually exclusive. 'Mental' is defined as qualitative, subjective, first personal, and therefore immaterial. 'Physical' is defined as quantitative, third personal, and therefore material... These definitions are inadequate to capture the fact that the world works in such a way that some biological processes are qualitative, subjective, and first personal. If we are going to keep this terminology at all, we need an expanded notion of the physical to allow for its intrinsic, subjective mental component... We do not live in several different, or even two different, worlds, a mental world and a physical world, a scientific world and a world of commonsense. Rather, there is just one world; it is the world we all live in, and we need to account for how we exist as a part of it."

Any 'one-world' approach must include the role played by human social/institutional constructs.

28. The ability of humans, by virtue of their thinking, choosing and acting, to affect what happens in the universe – albeit only on or near the tiny bit of it comprising planet Earth – appears indubitable. At a trivial level, our choices of daily activity determine the space-time location of the particles forming our bodies, clothes, cars, etc. At a less trivial level, theorising about the nature of reality at the micro-physical level has led to the acquisition by humans of the ability to trigger releases of nuclear energy – whether for peaceful or non-peaceful purposes – which would not otherwise occur and which potentially could destroy most, if not all, life on Earth. Crucially, how we choose to use the knowledge and expertise at our disposal depends upon the social/institutional reality³⁵ which we create collectively through the exercise of our imaginations. Included amongst our social/institutional constructs are the norms, codes of conduct and laws which provide the context for the attribution of personal responsibility – *itself* a social construct. A one-world approach (see previous paragraph) requires acceptance that the products of the imagination are as real and have as much causative power, potentially, as anything else. *What* is imagined, it is important to emphasise, does not have to be *true* in order to possess such a power. The racist *lies* promoted by Hitler's Nazi regime, for example, resulted in the mass murder of millions of human beings (see Appendix A) and, over millennia, the imagined wishes of imagined 'gods' have been invoked to justify the persecution and killing of anyone, including the adherents of rival religions/sects, not sharing prescribed beliefs. Similarly causatively powerful and equally dangerous are the myths we are prone to create regarding, for example, 'ethnicity' and 'nationality'. To write off all such phenomena as the blind product, ultimately, of elementary particles in fields of force is to abandon any attempt to render coherent the nature of human agency and the part it plays in shaping world events. Fundamental to any such attempt is recognition of the implications of *system complexity* and *emergence*.

Entities, from atoms to humans, may be viewed as systems. As systems combine together, more complex systems are produced. Rising complexity can result in the emergence of wholly new properties.

29. Any *system*, by definition, must have two or more interacting parts and may itself be part of – i.e. a sub-system within – a larger system. The more extensive the hierarchy of its sub-systems the more complex is any overall system. An example at the low end of the scale of complexity is a hydrogen atom conceived as comprising one electron (an elementary particle) and one proton (itself comprising three elementary particles – two up quarks and one down quark). At the high end of the scale is the human brain functioning through a complex network of parallel-processing sub-systems. A crucial feature, as system complexity rises, is the *emergence* of wholly *new properties* not displayed or suggested by the properties of lower level systems. As Gazzaniga (2011) states: "A complex system is composed of many different systems that interact and produce emergent properties that are greater than the sum of their parts and cannot be reduced to the properties of the constituent parts... Emergence is when micro-level complex systems... self organise... into new structures, with new properties that previously did not exist, to form a new level of organisation at the macro level." He points out that this is true of something as simple as the propensity of balls to roll down a slope. "The balls... are made up of atoms that behave as described by

³⁵ Explored by John Searle in *Making the Social World: The Structure of Human Civilization* (2010).

quantum mechanics, and when those microscopic atoms come together to form macroscopic balls, a new behaviour emerges and that behaviour is what Newton observed and described. It turns out that Newton's laws aren't fundamental, they are emergent; that is, they are what happens when quantum matter aggregates into macroscopic fluids and objects. It is a collective organizational phenomenon. The thing is, you can't predict Newton's laws from observing the behaviour of atoms, nor the behaviour of atoms from Newton's laws. New properties emerge that the precursors did not possess. This definitely throws a wrench into the reductionist's works and also throws a wrench into determinism."

In life forms which have evolved highly complex brains, a key emergent property is *consciousness*.

30. Crucial to the scope for *agency* is the emergence, through a process of evolution, of *consciousness* as a property of life forms with complexly structured brains – in particular, the advanced form of consciousness possessed by humans which enables them to function as rational and moral agents. Swaab (2014) – who cites, as a simple example of emergence, the wholly new properties which arise when oxygen and hydrogen atoms combine to form water – argues: "Consciousness can be seen as an emergent characteristic generated by the joint functioning of specific areas of the huge network of neurons in our heads. Brain cells and areas have their own separate functions, but their functional links with one another jointly endow them with a new 'emergent' function." Similarly, Gibb (2012) states: "Today, most scientists and philosophers agree that ... 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 astonishing that something so 'physical' as electro-biochemical processes within the brain could produce something so intangible as consciousness, but this is what happens. We just don't yet understand how."

The view of consciousness as a *high level* emergent property contrasts with that of panpsychism. Mental states appear capable of *constraining* the brain states which give rise to them.

31. Although 'physical substance' or 'matter' – conceived as the manifestation of elementary particles in fields of force – would appear to possess the *potential* for consciousness, the implication of emergence is that such potential is realised *only* when, through a process of system-building, it is structured in very specific and highly complex ways.³⁶ The view of consciousness as a *high level* emergent property thus differs from that of panpsychism (see paragraph 10) in that it does not envisage consciousness as featuring at all levels of complexity. Of crucial importance to the issue of free will and responsibility is the potential for properties which emerge at higher levels of complexity not just to differ from, and be unpredictable in terms of, lower level properties (see paragraph 30) but also to *constrain* lower level activity. Gazzaniga suggests that: "Mental states that emerge from our neural actions do constrain the very brain activity that gave rise to them. Mental states such as beliefs, thoughts, and desires all arise from brain activity and in turn can and do influence our decisions to act one way or another. Ultimately, these interactions will only be understood with a new vocabulary that captures the fact that two different layers of stuff are interacting in such a way that existing alone animates neither." Arguably, however, any new vocabulary has to be combined with a new way of *conceptualising* the reality involved. Conceiving mental/brain *activity* as 'layers of *stuff*' seems inappropriate and to involve a category-mistake (see paragraph 10). *Activity*, whether mental or physical, appears *flow-like*, *time-extended* and irreducible to a succession of instantaneous, and thus *duration-less*, 'states'.

The relationship between mental activity and brain activity might best be understood in terms of *complementarity* rather than *upward/downward causation*.

32. A problem with conceiving mental and brain activity in terms of successive 'states' is the sequence of 'upward' and 'downward' causations which then appear necessary for mental states to constrain brain states – e.g. for mental state M1 arising from brain state B1 at time T1 to moderate brain state B2 at time

³⁶ The process of evolution, from primordial slime to humans possessing consciousness as an emergent function, has taken over a billion years to complete. Now that the DNA 'blueprint' for a human being exists, however, the 'system-building' involved in the creation of each new one takes a mere nine months or so i.e. the time between human conception and birth.

T2 and thereby mental state M2 at time T2. Appendix C illustrates the process envisaged and how it diverges from the causal supervenience model (see paragraph 8) in which all mental activity is considered the causally inert product of causally effective but unthinking brain processes. *Both* models appear unsatisfactory. Gazzaniga suggests that the relationship between mental activity and brain activity might best be viewed not in terms of upward/downward causation (the idea of which appears to involve an inappropriate and misleading *spatial* metaphor) but of *complementarity* – i.e. the mutual interaction of separate but somehow related features of the one reality. A perhaps useful analogy is the relationship between computer software and hardware – the information supplied by the former being *realised through*, but at the same time *directing the functioning of*, the latter (the micro-physical components of which are not restricted to pre-determined or random paths but can *respond* to the instructions involved).

A new level of emergence arises when individual brains *interact*. Key properties which emerge at this level are *responsibility* and *freedom*. Each level of emergence has its own set of 'rules'.

33. Gazzaniga identifies a level of emergence *beyond* that of consciousness in individual brains – i.e. one which arises when those brains *interact*. It is this interaction which spawns the social constructs (see paragraph 28) which provide the framework for the attribution of responsibility for our acts and omissions. As Gazzaniga says: "Responsibility is a dimension of life that comes from social exchange, and social exchange requires more than one brain. When more than one brain interacts, new and unpredictable things begin to emerge, establishing a new set of rules. Two of the properties that are acquired in this new set of rules that weren't previously present are responsibility and freedom." It is at *this* level of emergence that we operate as agents in the world, accepting responsibility, within limits, for what we do, or fail to do. As long as our interventions are instrumental in achieving desired results, why worry about how they connect with lower level processes, fascinating though this might be from a scientific and philosophical point of view?³⁷ Properties emergent at a given level of complexity, in any case, have to be understood and manipulated on their own terms. We cannot decide what we *should* or *shouldn't* do in life through the application, for example, of Newtonian physics, Einsteinian relativity theory or quantum mechanics.

In practice, causal determinism does *not* provide an effective means of defence for wrongdoers.

34. Our social constructs reflect our view of the nature of reality – and this can vary between individuals. The belief that our every thought, word and deed is 'predestined from all eternity' is genuinely, if nebulously, held by some people – even though they *in practice* live their lives as if weren't true (see paragraph 6). As such belief may affect how they behave (see paragraph 16), it cannot be entirely ignored and needs to be confronted by the type of arguments contained in this paper. It does not appear, however, to hold much, if any, attraction for people who seek to defend themselves against accusations of wrongdoing. As illustrated in Appendix A, their line of defence is generally to deny the evidence against them or, if it proves irrefutable, to seek to justify what they did or shift responsibility to someone above them in a chain of command. If they deny responsibility on the grounds that they were *pre-determined* to do what they did, they invite the following response from their accusers: "We are *equally* pre-determined to do what *we* do, so it is pointless for you to complain if we now find you guilty and hang you". Causal determinists, unsurprisingly, will generally find themselves insisting that those intent on doing something very nasty to them are *not* pre-determined in their actions and *do* have the freedom to choose otherwise!

In practice, belief in predestination has raised fewer problems than belief in moral absolutes.

35. Historically, far fewer problems have arisen from the belief that all our actions are predetermined than from the belief that they are to be judged against moral rules which are *unarguable* either because they are prescribed by a supposed deity (for whom a self-appointed priesthood usually claims the exclusive right to speak) or because they are supposedly self-evident, being somehow intrinsic to 'the way things are

³⁷ Compare this with the view expressed by Russell (1925) that physicists do not need to understand the intrinsic nature of things (which may be essentially unknowable) in order to manipulate them by applying rules which *work* – i.e. produce desired results – at a given level of perception (see paragraph 22).

in themselves' – viz. Kant's categorical imperatives (see paragraph 25). Both approaches *deny* our freedom to work out for *ourselves*, combining rational and emotional argument, the norms by which we wish to live. Paradoxically, religious approaches (with some exceptions), whilst denying our freedom to choose those norms, insist that we *are* free to choose whether or not to *observe* them³⁸, our performance in this respect generally being linked to our prospect of eternal salvation or damnation.

Attributing responsibility to human agents should serve a *reformative* rather than *retributive* purpose.

36. Religious and other types of metaphysical dogmatism have brought much confusion to the essentially practical issue of the types of behaviour we wish to encourage or discourage and how we might reasonably do this. Notions such as those of 'sin' and 'damnation' shift attention from the practicalities of channelling human behaviour in positive directions – although we may argue about what these are – to passing judgement upon people *as an end in itself*, thereby encouraging a *retributive* approach where the treatment of wrongdoers is focussed upon making them suffer rather than reforming them. It may also encourage a culture where problems are deemed solved just by blaming someone for them. Where bad things result from human acts or omissions, however, to attribute blame – purely in the sense of to "assign responsibility for a fault or wrong"³⁹ – is exactly what we have to do in order to fairly and effectively target appropriate reformative/preventative measures (see paragraph A6 of Appendix A).

With personal responsibility goes the need for *positive* criticism – of *ourselves* as much as of others.

37. From the above, we may conclude that the attribution of personal responsibility is crucial to any attempt to improve human behaviour – including our *own*. Often, indeed, we are our own sternest critics and far more ready to forgive others than ourselves. The challenge is make self-criticism a *positive* force, creating the will to *change*, through relevant means, our future behaviour – in the belief that we *are* free, within limits, to do so. A danger with self-criticism is that it can turn into pointless *self-recrimination* and even *self-indulgence* where we adopt a *mea culpa* mentality, over-dramatising our failings and berating ourselves for them. Poet James Thomson (1834-82)⁴⁰ provides a telling example of this in a poem he wrote in 1869 : "Once in a saintly passion / I cried with desperate grief, / "O Lord, my heart is black with guile, / Of sinners I am chief." / Then stooped my guardian angel / And whispered from behind, / "Vanity, my little man, / You're nothing of the kind." *Positive* self-criticism requires a sense of proportion coupled with a focus upon rectifying mistakes where possible and determining how to avoid them in future.

Being free to make our own choices does not require us to be free from *all* constraints, some of which are *necessary* if those choices are to be coherent and effective.

38. Accepting responsibility for what we do or fail to do appears of far greater *practical* importance than speculating about whether or not we possess 'free will', the notion of which is conceptually ambiguous and arguably unintelligible (see paragraph 1). As Gazzaniga states: "The issue isn't whether or not we are 'free'. The issue is that there is no scientific reason not to hold people accountable and responsible." In order to be at all responsible for our behaviour, of course, we must have at least *some* freedom to choose to behave otherwise (see paragraph 14). Such freedom, however, is necessarily *circumscribed* and not to be confused with the chimera of an *unfettered* free will. Gazzaniga points out that there is much from which we do *not* want to be free when we make choices. "What does it really mean to talk about free will? 'Ah well, we want to be free to make our own decisions.' Yes but what do we want to be free from? We don't want to be free from our experience of life, we need that for our decisions. We don't want to be free from our temperament because that also guides our decisions. We actually don't

³⁸ The Catechism of the Catholic Church, for example, includes statements such as "Freedom is the power, rooted in reason and will, to act or not to act, to do this or that, and so to perform deliberate actions on one's own responsibility" and "Freedom characterizes properly human acts. It makes the human being responsible for acts of which he is the voluntary agent."

³⁹ Concise Oxford Dictionary (2001) - tenth (revised) edition.

⁴⁰ He should not be confused with the 18th century Scottish poet James Thomson (1700-48) who is best known for writing the words of *Rule Britannia* and who lived for the last twelve years of his life in Richmond upon Thames, spending many happy hours in Richmond Park (see Appendix D).

want to be free from causation, we use that for predictions... We don't want to be free from our successfully evolved decision-making device." Julian Baggini (2015), similarly, recognises that our choices and actions are bound to be influenced by our desires, beliefs, principles and personalities but that this does not make them un-free. If they were *not* so influenced, how they could be regarded in any sense as *our* choices and actions? Taking the freedom exercised by an artist as an example, he argues: "First of all... to be free is for your choices to flow from you, whether they are entirely conscious or not. Second, to be free is to be able to generate highly personal outputs from the inputs of nature, nurture and society, not to be free from their influences, able to create from nothing. Free choices are ones where the individual contributes something indispensable to the choice, even if the ability to make that contribution is something that is in one sense simply the result of nature and all past experience – for what else could it be the result of? Third, to be free is to make choices in the knowledge that there are other options and without being forced or coerced in one way or another. This can be the case even if, from a certain point of view, the choice you actually make is the only one you would ever have made in that situation."

Theories consistent with our *practical* experience of freedom and responsibility are more credible than those which contradict that experience and are wholly counterintuitive.

39. We should not expect to find any easy resolution of the conceptually challenging issues explored in this paper regarding our freedom to make choices for which we bear responsibility. However, some approaches appear much more helpful than others. The emergence of wholly new properties as system complexity rises is suggestive, at least, of how the relationship between mental and brain activity might be one of mutual *complementarity*, rather than the former being just the inert product of the latter (see paragraphs 31 & 32). At the level where highly complex sentient/cognitive systems (i.e. humans and their brains) *interact*, new 'rules of engagement' do appear to emerge including the attribution of responsibility for choices freely made (see paragraph 33). This is consistent with the way we *in practice* make choices in our lives (see paragraph 6), hold each other responsible for our behaviour and apply appropriate measures, including social/legal sanctions, to influence it (see paragraph 18). Theories which conclude the opposite – e.g. that our behaviour is determined *entirely* by forces beyond our control and that we are not responsible for anything we do – are counterintuitive and contradict our practical experience. Whilst this does not *disprove* them it must cause us to view them with extreme scepticism and to suspect the coherence of their 'reasonings'. In conclusion, it is important to emphasise that 'philosophy' – as an intellectual discipline which examines the conceptual armoury we use when trying to make sense of all that we experience – must address, and be relevant to, the *realities* of that experience. Rather than concoct an idealised world, it must attend to the actual, if somewhat messy, world we inhabit – a world in which we, as highly complex *cognitive systems* possessing *agency*, regularly make choices for which we accept responsibility, unable to hide behind excuses such as 'the universe, my brain or sub-atomic particles made me do it'!

Roger Jennings
December 2019

Main texts referred to in this paper:

- Baggini, Julian (2015) *Freedom Regained: The Possibility of Free Will*. Granta
- Berkeley, George (1710) *The Principles of Human Knowledge*.
- Dennett, Daniel C. (1991) *Consciousness Explained*. Penguin
- Gazzaniga, Michael S. (2011) *Who's in Charge? - Free Will and the Science of the Brain*. HarperCollins
- Gibb, Barry J. (2012) *The Rough Guide to the Brain*. Rough Guides
- Heisenberg, Werner (1962) *Physics and Philosophy: The Revolution in Modern Science*. Penguin Classics reprint, 2000
- Kant, Immanuel (1781) *The Critique of Pure Reason*.
- Locke, John (1689) *An Essay concerning Human Understanding*.
- O'Grady, Jane (2018) *Enlightenment Philosophy in a Nutshell*. Arcturus
- Pink, Thomas (2004) *Free Will: A Very Short Introduction*. Oxford University Press
- Schopenhauer, Arthur (1818) *The World as Will and Representation*
- 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
- Solomon, Robert C. (1988) *Continental Philosophy since 1745: The Rise and Fall of the Self*. OUP
- Steward, Helen (2012) *A Metaphysics for Freedom*. Oxford University Press
- Swaab, Dick (2014) *We Are Our Brains: From the Womb to Alzheimer's*. Penguin Books

A paper entitled *Freedom of Choice and Action* was considered at the Kingston Philosophy Café session held on 30 September 2015. The text of the paper can be accessed via the following link:

https://e-voice.org.uk/kingstonphilosophycafe/files/view/philosophy-cafe-briefings/paers-by-roger-jennings/Freedom_of_Choice_and_Action.pdf

Appendices:

	Pages:
A. Can we escape 'ultimate' responsibility for what we do?	21 - 25
B. Jean-Paul Sartre's <i>Roads to Freedom</i> trilogy.	26
C. The relationship between mental states and brain states.	27 - 28
D. A bit of related local history.	29 - 30

Appendix A: Can we escape 'ultimate' responsibility for what we do?

According to Galen Strawson, Professor of Philosophy at the University of Texas, nobody can be *ultimately* responsible for *anything* they do.⁴¹ If true, this would mean that none of the men shown below could be deemed ultimately responsible for the prominent role they played in the perpetration of *mass murder*. Can this possibly be right?



Adolf Hitler

Estimated number of civilians and disarmed soldiers killed by Nazi regime and its collaborators:

	millions
Jews (excluding Jewish Soviet POWs)	6.0
Non-Jewish Soviet civilians	5.7
Non-Jewish Polish civilians	1.8
Soviet POWs	3.0
Serb civilians	0.3
People with disabilities	0.3
Roma	0.2
Repeat criminal offenders and 'asocials'	0.1

The above figures are rounded and inevitably approximate. They total over 17 million people and exclude an unknown number of political opponents and resistance activists in Axis occupied territory.



Heinrich Himmler



Reinhard Heydrich



Auschwitz-Birkenau concentration/death camp where almost one million Jews are estimated to have been killed.



Adolf Eichmann



Radovan Karadžić



During the 1992-95 Bosnian War, an estimated 8,300 Bosniaks, mainly men and boys, were massacred in July 1995 at Srebrenica by Republika Srpska forces commanded by Ratko Mladić, himself responsible to Republika Srpska's president, Radovan Karadžić. The remains of 6,500 of the victims are buried at the Srebrenica–Potočari Memorial and Cemetery, shown above.



Ratko Mladić

⁴¹ See: <https://philosophy.as.uky.edu/sites/default/files/The%20Impossibility%20of%20Moral%20Responsibility%20-%20Galen%20Strawson.pdf>

A1. Strawson's basic argument against attributing ultimate responsibility to anyone for anything they might do, runs as follows.

1. You do what you do, in any given situation, because of the way you are.
2. To be ultimately responsible for what you do, you have to be ultimately responsible for the way you are – at least in certain crucial mental respects.
3. But you cannot be ultimately responsible for the way you are in any respect at all.
4. So you cannot be ultimately responsible for what you do.

A2. Strawson's argument is presented in the form of three premises (statements 1-3) and a conclusion (statement 4). The *truth*, as opposed to *logical validity*, of the conclusion depends upon the truth of the premises. A problem with identifying their truth is their *ambiguity*. The meaning to be attached to the word 'ultimately' is particularly obscure. In *practice*, ultimate responsibility for events tends to be attributed only to people who, by virtue of heading a chain of command, *initiate* the process resulting in their occurrence.⁴² Strawson is clearly *not* using the word in this sense but leaves unclear its intended meaning and, in particular, how being 'ultimately responsible' for anything might differ from being just plain 'responsible'.

A3. In Strawson's argument, the intended meaning of 'the way you are' is as unclear as that of 'ultimately'. Between birth and death, 'the way we are', both mentally and physically, is subject to *change* – usually gradual but occasionally dramatic. How we are *now* is substantially different from how we were, say, five, ten or twenty years ago. Apart from long-term change, we are also subject to short-term variations in behaviour – due, perhaps, to mood swings which may cause us sometimes to act 'out of character'. In Strawson's first premise, therefore, 'the way you are' would seem to have to refer to the *particular* physical/mental/emotional state we are in when, in a given situation, we do something. But we are then drawn into a *tautological* 'explanation' of why we do what we do i.e. that we do X in situation S at time T just because the way we are at time T happens to be such as to cause us to do X in situation S. The nature of the connection between the way we are at any time and what we do is left entirely obscure. Arguably, the distinction *itself* is incoherent. What we do – i.e. how we *behave* – is *part* of, not separate from, the way we are. Similarly there is no identifiable 'we' distinct from 'the way we are' – i.e. distinct from our total mental and physical make-up. Take this away and what remains but a *void*? So when Strawson says 'you cannot be ultimately responsible for the way you are', what is the 'you' to which he refers? Does he, perhaps, imagine some featureless 'ghost in the machine'?

A4. There seems no logical or practical reason why a human 'self' – conceived *not* as a 'ghost in the machine' but as a sensory/cognitive *system* possessing consciousness, intentionality and agency – should be unable to exercise at least *some* control over 'the way it is'. We may be stuck with our genes and much else besides but our everyday experience is nevertheless one of *choosing*, and thus being *responsible for*, what we think, say and physically do – themselves key aspects of 'the way we are'. If told by Strawson or anyone else that we cannot be *ultimately* responsible for the way we are and thus for what we do, we might suspect that the supposed impossibility arises not from anything about ourselves but from the *meaninglessness* of the notion itself. As suggested above, it has meaning *only* in relation to the apportionment of responsibility between people who, at different levels in a hierarchy, play a part in causing something to happen. Only intelligent beings can understand and judge the consequences of their behaviour and thus be held responsible for it. We cannot free ourselves of responsibility for what we do or fail to do by shifting 'ultimate' responsibility to unthinking things – e.g. 'upwards' to the entire universe or 'downwards' to elementary particles and waves.⁴³ The level at which we can be, and conventionally *are*, held responsible for our behaviour is as 'ultimate' as it gets.

⁴² In the first/inconclusive trial held in 2019 in relation to the 1989 Hillsborough Stadium disaster where 96 people died, for example, specific reference was made to establishing who bore *ultimate* responsibility for what happened.

⁴³ Believers in a Creator Deity might attribute to her/him/it ultimate responsibility for everything which happens in the universe. Such a being, of course, might then deploy the 'Strawson Defence' and disclaim ultimate responsibility for the way she/he/it is!

A5. It is possible, when Strawson refers to 'the way we are', that he has in mind some notion akin to that of 'facticity' – i.e. the experiential 'baggage' we continually accumulate, described by Solomon (1988) as "those deeds and events that are over and done with, but whose consequences largely determine our present circumstances, and constitute a significant part of who or what we are". If so, Strawson appears to assume, without argument or justification, that the content of our facticity at any moment doesn't just *influence* what we do at that moment but absolutely *determines* it. But why assume this? Why rule out the possibility that 'the way we are' is such as to allow us to *transcend* our facticity and *originate* our actions in ways which are *non-deterministic* (see paragraph 15)? If Strawson espouses some form of *determinism*⁴⁴ or of *causal supervenience* (see paragraphs 7 & 8) he needs to spell it out and justify it.

A6. Consideration of our responsibility for what we *do* requires clarity as to what *doing* something involves and how issues of responsibility might arise in connection with it. The following points are relevant.

- a) To engage in mental activity – e.g. thinking, imagining, day-dreaming, planning and intending – is to *do* something even if not linked to physical activity. In some societies, to merely *hold* a proscribed religious or political belief is to do enough to merit persecution, imprisonment and even death.
- b) Often the *physical* component of what we do is minimal and only a means rather than an end in itself. An obvious example is our intercommunication via the spoken/written word and other signs. The physical processes involved are themselves trivial, their significance deriving purely from the *meaning* they convey.
- c) It is the *intentional* content of our actions – whether or not they have a significant *physical* element – which gives them meaning. *Identical* physical acts may be interpreted differently if the intentions behind them differ – e.g. if I hit someone hard enough to kill them, it is a judgement about my intentional state at the time and in the circumstances concerned, which will determine whether what I do is deemed to be murder, manslaughter or justified self-defence.
- d) Much of what we do in life has meaning only within the context of the *social/institutional reality* which we create and sustain through the collective exercise of our imaginations. The *mental constructs* involved are wide-ranging and include religions, nation states, governments, companies, property, money and marriage. The content of, and practices associated with, such constructs may vary substantially between human groups and, indeed, provide the basis for their differentiation e.g. groups distinguished by nationality, religion or class.
- e) Holding people responsible for their intentional acts is not an 'either-or' exercise. In practice we distinguish *degrees* of responsibility depending, in particular, upon the nature of any constraints upon their freedom of choice, their temporary or long-term capacity to make informed and reasoned judgements and the significance of their role relative to that of others (see paragraph 14).
- f) We may be held responsible for *unintentional* acts – e.g. inadvertently exceeding speed limits – if it is considered we *could* and *should* have avoided them by exercising better control of our behaviour. There is thus a *normative* element to the attribution of responsibility predicated upon the belief that we *can*, within limits, choose/control/moderate how we behave.
- g) We are commonly held responsible not just for what we *do* but also for what we *fail* to do, whether intentionally (e.g. deliberately ignoring safety regulations) or unintentionally (e.g. forgetting to keep an appointment). Again, such responsibility arises only by virtue of normative judgements about how we could and should behave.
- h) The proper attribution of responsibility to individuals for their acts and omissions is essential if legal measures (such as fines, community service and imprisonment) designed to minimise unwanted behaviour are to be fairly and effectively targeted. What generally needs to be established is not only a

⁴⁴ Determinism can be broadly defined as "the doctrine that all events and actions are ultimately determined by causes regarded as external to the will" – *Concise Oxford Dictionary*, Tenth (revised) Edition, 2001.

causative connection between individuals and identified acts/omissions but also that they might have chosen to behave otherwise.⁴⁵

The above points serve to highlight the complexity of issues surrounding the attribution of personal responsibility, including its *normative* aspects and the crucial importance of the *mental/intentional* states of the human agents involved.

A7. It has to be said that Strawson's 'proof' of the impossibility of anyone being ultimately responsible for anything they do smacks more of *sophistry*, in pseudo-logical guise, than philosophy and owes any superficial plausibility it might have to the ambiguity of its language. Strawson, presumably, would deny that he is 'ultimately' responsible for formulating it – or, indeed, for anything else he might do! Particularly striking is its *practical irrelevance*. As already pointed out, it does not purport to disprove the possibility of our being *responsible* for what we do or fail to do – only, and meaninglessly, our being *ultimately* responsible. It treats the attribution of responsibility, moreover, as a purely descriptive matter, ignoring its essential *normative* element. There is evidence that such attribution, linked to judgements about the degree to which individuals are free to make informed and reasoned choices, has the power to influence how we behave (see paragraph 16) and thus has *instrumental value*. Attributing responsibility, as appropriate, to ourselves and others for what we do or fail to do is thus *itself* something for which we carry responsibility. What is deemed appropriate will depend upon how we apply the type of criteria outlined in paragraph 14. It will also be affected by our *view of reality* – in particular, what causative relationships we believe are possible between events and human agents. In societies which entertain primitive superstitions about 'witchcraft' and 'devilry', for example, it can mean *death* for unfortunate individuals deemed responsible for plagues, crop failures, floods, etc.

A8. If Strawson's argument were accepted as intelligible and true, what would be its *practical* implications for the treatment of individuals (such as the six men shown on page 21) proved to have been involved (whether by issuing orders, acting in an administrative capacity or supervising/doing the actual killing) in the perpetration of mass murder? What weight would their supposed lack of *ultimate* responsibility for their actions carry? In reality, three of the men never faced trial (Heydrich was killed by the Czech resistance in 1942 and Hitler and Himmler committed suicide in 1945). Had they survived and been tried at Nuremberg, they undoubtedly would have been found guilty and executed. It is hard to believe that the verdict would have been any different had they pleaded in their defence that: a) they did what they did because of the way they were; b) they could not be ultimately responsible for the way they were; c) therefore they could not be ultimately responsible for what they did. Unsurprisingly, this defence was not offered by Eichmann, Karadžić or Mladić when tried for the part they played in the commission of mass murder and other atrocities.

A9. Eichmann, who organised the transportation of victims to Nazi death camps during World War II, was tried in Israel in 1961, charged with crimes against humanity, war crimes, crimes against the Jewish people, and membership of a criminal organisation. He was found guilty on 15 counts (although not of *personally* killing anyone or of controlling the activities of the *Einsatzgruppen* – nominally 'task forces' but in practice death squads) and sentenced to execution by hanging, the sentence being carried out in 1962 after a failed appeal. Eichmann initially sought to deny what he had done but, in the face of clear evidence, later sought to shift responsibility up the chain of command. In his plea he stated: "There is a need to draw a line between the leaders responsible and the people like me forced to serve as mere instruments in the hands of the leaders. I was not a responsible leader, and as such do not feel myself guilty." Eichmann

⁴⁵ This does not mean, of course, that people deemed incapable of rational/moral choice due to mental disorder are exempt from action under the law if they commit criminal acts, only that the nature of that action is liable to differ e.g. in serious cases it may involve confinement in a secure mental institution rather than a prison (viz. the Kray twins). Unlike the vast majority of people who can be targeted by the law only for what they *do*, not for what it is feared they *might* do, people with mental disorders may be 'sanctioned' under the Mental Health Act (i.e. detained/treated in a hospital, with or without their consent) if it is considered that they are likely to do something harmful to the health or safety of themselves and/or others.

appeared not to appreciate the enormity of his crimes and showed no remorse. He seemed to regard himself as a mere bureaucrat who was 'only obeying orders'. In her book *Eichmann in Jerusalem: A Report on the Banality of Evil* (1963), philosopher Hannah Arendt highlights the fact that people who, like Eichmann, are in many ways quite 'ordinary', may nevertheless do appalling things. In his trial Eichmann, who came over as a not very intelligent man, claimed always to have tried to observe Kant's categorical imperative that we should invariably act in accordance with universalizable rules – unfortunately interpreting these as the rules dictated by Hitler rather than those arising from the exercise, for which we each bear individual responsibility, of our own moral intelligence.

A10. Karadžić was indicted in 1996 by the UN International Criminal Tribunal for the former Yugoslavia (ICTY) for war crimes – including genocide against Bosniak and Croat civilians – committed during the Bosnian War (1992-95). He hid under a false identity but was eventually arrested and brought to trial in The Hague in 2008. He claimed to be the victim of a conspiracy and refused to accept the authority of the court, so a plea of not guilty was entered on his behalf. In 2016 he was found guilty on 10 counts relating to the genocide at Srebrenica, war crimes and crimes against humanity and sentenced to 40 years' imprisonment. His appeal against sentence was rejected in 2019 and the sentence was increased to life imprisonment. One piece of evidence against him relating to the Srebrenica massacre was an order he had signed four months earlier calling for conditions for the city's people to be made "unbearable with no hope of further survival". Mladić, like Karadžić, was indicted for war crimes by the ICTY and later went into hiding. He was located and arrested in 2011 and extradited to The Hague. His trial ran from 2012 to 2017 when he was found guilty on 10 counts of committing war crimes, crimes against humanity and genocide. He was sentenced to life imprisonment. As the top military officer with command responsibility he was deemed by the ICTY to be responsible for both the siege of Sarajevo and the Srebrenica massacre. His appeal against sentence is currently (December 2019) ongoing.

A11. Karadžić and Mladić defended themselves by refusing to accept the charges and evidence brought against them. Eichmann, initially, did the same but later sought to shift responsibility from himself to the 'leaders' whom he claimed he had no choice but to obey. None saw fit to argue that they could not be 'ultimately responsible' for anything they did because they could not be ultimately responsible for 'the way they were'. If they had, the argument would almost certainly have been rejected as essentially vacuous and irrelevant to the issue of their *actual* responsibility for what they were proved to have done. Holding people responsible for what they do involves identifying the following.

- a) Which things we consider *good* and to be *encouraged* or *bad* and to be *discouraged*.
- b) The *factual* contribution (i.e. its nature and extent) of individuals to the occurrence of such things.
- c) The ability of those individuals to make informed and reasoned judgements.
- d) The existence of any constraints upon their freedom of choice.
- e) What, if anything, should be done to *them* to control/influence their future behaviour and, by their example, that of *others* – on the presumption that people can, within limits, choose how they behave.

Naive determinists might argue with regard to c), d) and e) that people are *totally* constrained in their actions by forces beyond their control and that human choice is thus an illusion. It does not follow from this, however, that no action should be taken against those who do bad things. Strictly speaking, indeed, it means we can do whatever we like to them – including *kill* them – on the basis that, by the same token, anything *we* do must be equally 'predestined from all eternity' (see quote 2). Fortunately for offenders, we do not take this line and believe that *we can* choose, and are thus responsible for, what we do. Our moral precepts are not fixed but *evolve* as our moral sensibilities are refined through the exercise of our imaginations. This is clearly evidenced in our changed attitude towards capital punishment. Had Karadžić and Mladić committed their crimes some years earlier, they might, like Eichmann and ten of those sentenced at Nuremberg for Nazi war crimes,⁴⁶ have received the death penalty and been executed.

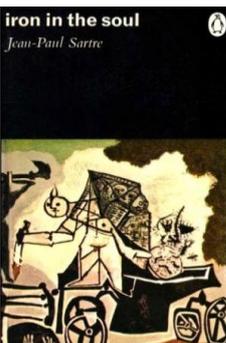
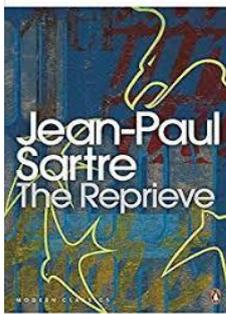
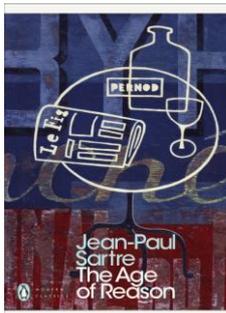
⁴⁶ An eleventh, Hermann Goering, was sentenced to death but avoided hanging by swallowing a cyanide pill the night before.

Appendix B: Jean-Paul Sartre's *Roads to Freedom* trilogy.

The major and best known philosophical work of Jean-Paul Sartre (1905-80) is *Being and Nothingness* (1943) but he also explores his 'existentialist' approach through the medium of fiction, most notably in his *Roads to Freedom* trilogy comprising, as generally translated from French into English, *The Age of Reason* (1945), *The Reprieve* (1945) and *Iron in the Soul* (1949).

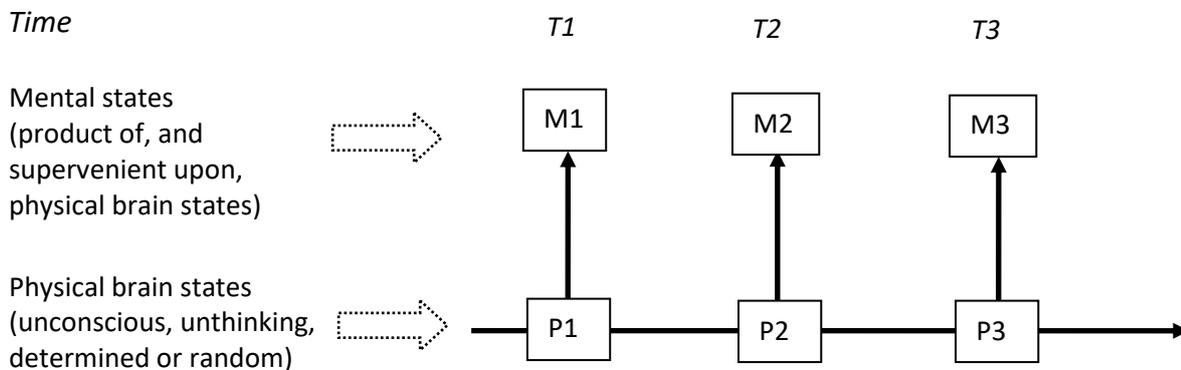
According to Sartre, an atheist, the absence of a Creator means that there is nothing pre-determined or 'essential' about human nature. It is up to humans themselves, once they have come into existence, to determine what they *become* – an idea encapsulated in his aphorism "existence precedes essence". If we fail to use our powers of self-determination – which includes the power to 'transcend' the influence of our accumulated life experience or 'facticity' (see paragraph 15) – we deny ourselves the opportunity to live 'authentically' and are guilty of 'bad faith'.

The *Roads to Freedom* trilogy is set in France in the lead up to, and at the start of, the Second World War and focuses upon Mathieu Delarue, a philosophy teacher and socialist, and his circle of friends. At the end of *Iron in the Soul*, Mathieu, now a soldier in the French army as the Germans invade France, has the choice, with a few others, of making a futile, and likely fatal, stand in defence of a French village or abandoning it to the enemy. His decision to stay and fight represents an ultimate act of transcendence, an affirmation of his freedom to live authentically and determine what he becomes. Sartre himself was drafted into the French army at the start of the War, was captured and spent nine months in a prisoner-of-war camp – where he read and was much influenced by *Being and Time* (1927), the major work of German existentialist philosopher (and Nazi Party member!) Martin Heidegger (1889-1976).

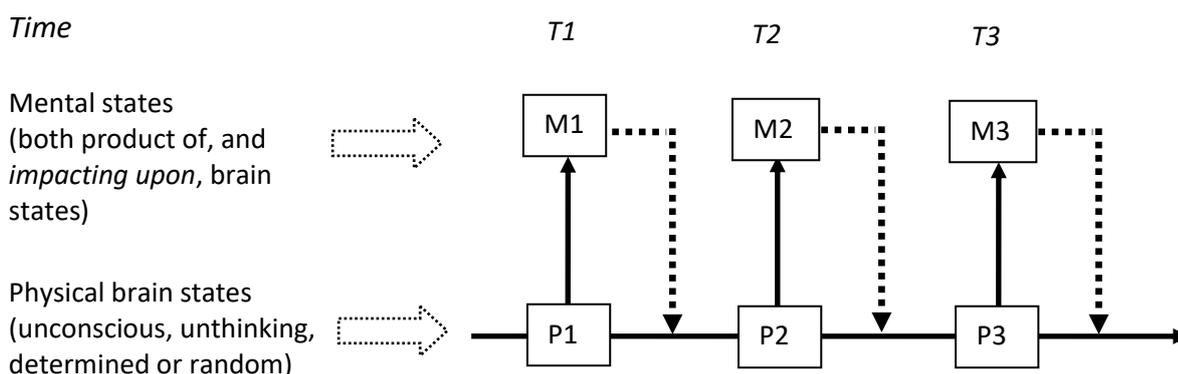


Appendix C: The relationship between mental states and brain states.

One view of the relationship between mental states and brain states is that of *causal supervenience* (see paragraphs 7 & 8 of this paper). Each mental state is viewed as the *inert* product of a brain state and thus *itself* incapable of influencing the mental state which follows. Only brain states are deemed to be causally connected to each other (although the possibility of some occurring randomly is not excluded). The diagram below provides a simplified representation of the relationship envisaged.



Viewing our thoughts as just a supervenient froth on the surface of *unthinking* brain and lower level activity is, to say the least, counterintuitive. We naturally regard our thoughts as *themselves* causally effective, one typically leading to another. Might the causal supervenience model be revised to incorporate *downward* causation from the mental to the physical? Gazzaniga (2011) explores this possibility as follows. "Once a mental state exists, is there downward causation? Can a thought constrain the very brain that produced it? Does the whole constrain its parts? This is the million-dollar question in this business [i.e. neuroscience]. The classic puzzle is usually put this way. There is a physical state, P1, at time 1, which produces a mental state, M1. Then after a bit of time, now time 2, there is another physical state, P2, which produces another mental state, M2. How do we get from M1 to M2? This is the conundrum. We know that mental states are produced from processes in the brain so that M1 does not directly generate M2 without involving the brain. If we just go from P1 to P2 then to M2, our mental life is doing no work and we are truly just along for the ride. No one really likes that notion. The tough question is, does M1, in some downward-constraining process, guide P2, thus affecting M2?" To represent this possibility, our causal supervenience diagram can be revised as follows.



The diagram shows each mental state 'getting in first' before the physical state from which it arose can *absolutely* determine the physical state which follows – e.g. M1, arising from P1, has a causative effect upon the transition from P1 to P2 and thus, via P2, the content of M2. The mere adding of lines to a diagram, of course, does not *explain* the relationship between mental states and brain states nor, more generally, between our *perceptual/cognitive experience* and the *brain/microphysical activity* associated with it, such activity – along with everything else we categorise as 'physical' – being *known* to us only by virtue of such mental experience and in terms of the concepts it employs. *Both* of the conceptual models

shown above appear unsatisfactory, particularly in their representation of the relationship between mental and physical *activity* in terms of spatially separated, higher and lower, 'states'. Representing essentially *flow-like* phenomena in semi-static terms – i.e. as sequences of *separate* but, in some unexplained way, *connected* instantaneous and thus *duration-less* 'states of affairs' – appears not just inappropriate but incoherent. Viewing them instead in terms of distinct 'events' or 'groups of events', as suggested by Bertrand Russell (see paragraph 22 of this paper), appears no more helpful. 'Boundary' problems arise concerning the sub-division of *continuums* of activity, whether this is into discrete 'events' or 'states'. We are clearly at the limits of our understanding – which is unsurprising with any attempt to 'think about thought' and how our thoughts relate to their *objects*, which include our own brains.

Appendix D: A bit of related local history.

Pembroke Lodge in Richmond Park was, from 1876 to 1894, the childhood home of Bertrand Russell – mathematician, logician, philosopher and social/political commentator and activist. He was born in 1872 into a long-established aristocratic family (he later became the 3rd Earl Russell). Orphaned at the age of three, he and his elder brother were taken into the care of his grandfather Lord John Russell (British Prime Minister 1846-52 and 1865-66) who lived at Pembroke Lodge. When his grandfather died in 1878, his religiously conservative but socially liberal grandmother took on sole responsibility for his upbringing. In his autobiography, Russell describes how at Pembroke Lodge he "grew accustomed to wide horizons and to an unimpeded view of the sunset". Russell died in 1970 at the age of 97. His best known works include *The Principles of Mathematics* (1903), *Principia Mathematica* (with Alfred North Whitehead) (1910-13), *The Problems of Philosophy* (1912) and *A History of Western Philosophy* (1945).



1858

Pembroke Lodge



Now

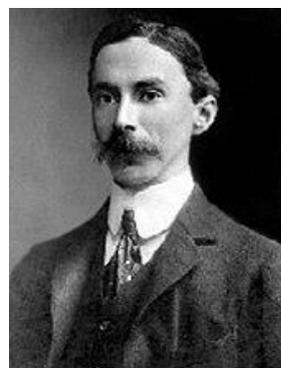
View from nearby Richmond Terrace



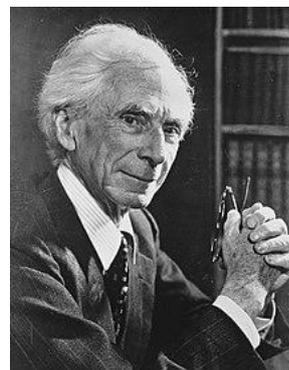
Bertrand Russell



1876



1907

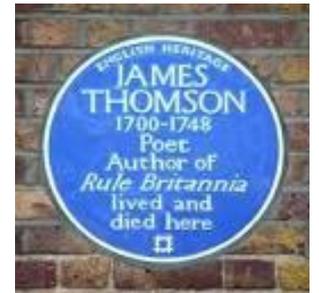


1950s

Richmond Park has an association with the 18th century poet and dramatist James Thomson (1700-48). He was born, raised and educated in Scotland, studying metaphysics, logic, ethics and natural philosophy (physical science) at Edinburgh University. He moved to London in his mid 20s and for the last twelve years of his life lived in Richmond upon Thames. His house in Kew Foot Lane is now incorporated into Richmond Royal Hospital and bears a commemorative plaque. He is best known for his poem *The Seasons* and for the words of *Rule Britannia* (set to music by Thomas Arne) which he is said to have written in *The Dove*, a riverside pub in Hammersmith (close to William Morris' Kelmscott House). He died at the age of 47 from a chill contracted on a river journey between Hammersmith and Kew and is buried in St. Mary Magdalene Church, Richmond. He spent many happy hours in Richmond Park where a board in the grounds of Pembroke Lodge commemorates his great love of nature with lines which begin:

"Ye who from London's smoke and turmoil fly, / To seek a purer air and brighter sky, / Think of the Bard who dwelt in yonder dell / Who sang so sweetly what he loved so well..."

There is a memorial to Thomson (next to Shakespeare's) in Poets' Corner, Westminster Abbey.



The poem quoted in paragraph 37 is the work of *another* James Thomson (1834-82), sometimes confused with his more illustrious namesake being also born in Scotland and also dying at the age of 47. From the age of seven, he was raised in a London orphanage. He served in the army for over a decade and, when stationed in Ireland, fell under the influence of Charles Bradlaugh who had gained notoriety as a free-thinker and atheist. After leaving the army, Thomson sought to make a living by writing stories, essays and poems – using the pseudonym Byshe Vanolis. For much of his adult life he struggled with insomnia, alcoholism and depression – perhaps brought on initially by the death, when he was a young man, of a woman with whom he was deeply in love.

In his poem *Philosophy* (1866) Thomson expresses his concern with the implications of rational/scientific thought for the prospect of finding meaning in life, describing how his central character *"Looked through and through the specious earth and skies"*. He strikes, nevertheless, an optimistic note when he argues for the importance of human love. In his most famous poem *The City of Dreadful Night* (1874), Thomson's mood is at its darkest and most pessimistic as he expresses his revulsion at the dehumanizing, uncaring environment of Victorian London – a flavour of which is provided in the following lines: *The City is of Night; perchance of Death / But certainly of Night; for never there / Can come the lucid morning's fragrant breath / After the dewy dawning's cold grey air: / The moon and stars may shine with scorn or pity / The sun has never visited that city, / For it dissolveth in the daylight fair."* Sadly, he did not manage to escape his depression and his darker thoughts by flying from *'London's smoke and turmoil... to seek a purer air and brighter sky'* in Richmond Park – the place much loved by his 18th century namesake.