Freedom and Thought

Michael Bourke

Abstract

Despite recent neuroscientific research purporting to reveal that free will is an illusion, this paper will argue that agency is an inescapable feature of rationality and thought. My aim will not be to address the methodology or interpretation of such research, which I will only mention in passing. Rather, I will examine a collection of basic concepts which are presupposed by thought, and propose that these concepts are interrelated in ways that makes them both basic and irreducibly complex. The collection includes such concepts as belief, value, meaning, and truth. I will argue that free will belongs to this collection, and as such is also presupposed by thought. This proposal is opposed to a methodological tendency in analytic philosophy, to eliminate aspects of concepts which can't be given a clear analysis, and to the wish of many empirical psychologists and cognitive scientists to reduce intentional/mental states to neurons and other mindless phenomena which they regard as more fundamental. Instead of offering a direct critique of either of these methodological attitudes, I will try to place the concept of freedom in its proper conceptual context, and make a positive case for its reality.

1. Language, thought and the self

Like many metaphysical concepts, free will has an uncertain range of application, so uncertain that it's not clear it has any. Perhaps few of us would say that conscious animals incapable of self-conscious deliberation possess free will. Many philosophers would allow that rational, language-using beings who ponder and revise the basis of their beliefs and values can sometimes freely act. But recently reports from neuroscientific research claiming that free will is an "illusory afterthought" of unconscious brain activity have been broadly circulated.[1] If this research is borne out, we should not be surprised if such claims become an accepted dogma among the general public, if they haven't already. The significance of our social institutions and forms of life that define our identity as persons presuppose belief in free will. But then the significance we assign to these phenomena may also come to be regarded as illusory. Objectively, it isn't obvious that freedom in a positive sense actually exists, even if we commonly feel that we experience our agency, more or less continually, through faltering and successful decisions. Human and non-human animals alike may simply be stimulus-response mechanisms which respond to the demands of the situations in which they find themselves exclusively on the basis of genetic dispositions and habits and beliefs instilled by interaction with their environment. If this familiar characterisation accurately describes our condition, the subjective experience of freedom in the overlapping contexts of our aesthetic, ethical, practical and cognitive existence may be an illusion. Free will may either be a pseudo concept, unintelligible in principle, or a concept that does not properly apply to creatures like us but only

to fantastical beings whose free existence we relegate to fantasy. The practical personal and social cost of this sceptical conclusion may be high; but perhaps it could be contained by policies that encourage widespread hypocrisy, double think, consumeristic diversion, intellectual apathy, resilient dogmatism, sophistry, and so on. Whatever the practical cost, the conceptual cost of eliminating the concept of free will, I will argue, is unthinkable. The view that I will try to make plausible is this: that we can't accept the conceptual cost of eliminating free will, since we would also need to eliminate a collection of interrelated cognitive and intentional concepts on which thought fundamentally depends.

Section two begins with a discussion of several cognitive concepts which underlie free will and might seem to be more basic, but which, I will propose, are inseparable from it. These underlying concepts are subsumed by the concepts of the self, language and thought, which in turn presuppose these concepts. The concept of free will is similarly related to the self, language and thought. It has no application without selves, subjects who exist more or less in the ways that we commonly seem to manifest ourselves to ourselves and to others, in intimate interactions and over wider more public spheres. The self in this common, manifest sense, which includes our experience of freedom, is assumed by the language with which we speak and think about reality. Language, in turn, taken as an activity in which meaning is communicated and understood, presupposes the thought, consciousness, and agency of individual language users. This very brief initial sketch is meant to suggest that self, language, thought and agency are interrelated concepts, each dependent on the others.

Perhaps all these concepts, related or not, are mere abstractions, which help us organise the things and relations of our ontology but themselves refer to nothing actual in the world. For instance, one possibility, they may refer only to virtual concepts, an implication of one of the three strands of Daniel Dennett's pragmatic, epistemic-stance approach to reality (or near reality). By adopting the *intentional stance* – which includes such items as selves, beliefs, thoughts, opinions, wishes, points of view, and so forth – we are able, on Dennett's view, to communicate with each other and to negotiate our way in the world by talking and acting *as though* selves, thoughts, beliefs etc. actually exist; we can organise our experience in ways that help us predict events, while realising that the assumed concepts of the intentional stance have no ontological significance.[2] If it turns out that free will also proves to be helpful in letting us talk predictively about our activities and practices, it too might be granted the status of *virtual entity*, or *virtual relationship*, or *virtual activity*.

Compatibilist attempts to save free will (the self, belief, etc.) by invoking strategies that encourage us to fill the world with "semi-real," virtual entities, to give a virtual status to entities that would otherwise be eliminated, may seem alluring if the alternative is simply to forsake talk of such thing. But such strategies, however well-intentioned and carefully articulated, skirt

sophistry. The reluctance to see basic concepts of thought more straightforwardly as integral to real-world relations and activities is misconceived. Donald Davidson offered a more promising approach to the role that mental-intentional concepts play in our language and ontology by offering the view that many of these concepts are *presupposed* by thought. Davidson was prompted to assert this view while responding on one occasion to Richard Rorty's criticism that truth is not the only concept that we require to describe and explain "human behaviour." Here is Davidson's response: "[W]hy, [Rorty asks,] is truth more important than such concepts as intention, belief, desire, and so on."[3] Davidson agreed that there are many other (closely and distantly) related concepts that we rely on. But his response extends and modifies Rorty's objection:

> Importance is a hard thing to argue about. *All* these concepts (and more) are essential to thought, and cannot be reduced to anything simpler and more fundamental. Why be niggardly when awarding prizes; I'm happy to award golden apples all round.[4]

Davidson extends Rorty's point when he says that "truth . . . belief, desire, and so on" are *fundamental* conceptual presuppositions, not merely pragmatic props. More importantly, he shifts the point from the issue of "human behaviour" to the related issue of *thought*. We need, at innumerable points, to refer to human behaviour and other parts of observable or extrapolated reality if we wish to understand concepts which are presupposed as fundamental to thought; but if the collection of concepts is meant to ground the possibility of *thought*, presumably it must include concepts within language that refer to subjective, intentional kinds of things and relationships, in addition to objective things and relationships. We might ask, though, 'Why can't the allegedly fundamental intentional concepts of thought "be reduced to anything simpler"'?

2. Irreducible complexity

Let us suppose that the presuppositions of thought encompass this (incomplete) collection: truth, meaning, value, belief and desire. The collection is unwieldy. It includes concepts that appear to refer exclusively to abstract entities, e.g. truth, and at least one concept, desire, which seems to refer to something reducible to particular physical processes – but not quite, as desire is also intentional and as such has a content and purpose, or, if half-formed, lists in the direction of content and purpose. Meaning on the other hand, like truth, with which it is entangled, seems to refer to something purely abstract. By contrast, belief appears to be more like desire, in that it is neither reducible to physical processes, nor to its abstract content – its propositional meaning. Perhaps value can be conceived several ways: e.g. abstractly as a type of meaningful statement; or, to suggest a less abstract connection with truth, as a disposition embedded in the truth commitment of every belief; or, more modestly, as roughly equivalent to an individual's desire.

To gain a clearer, more distinct view of these underlying concepts we might try disentangling them. In the case of desire and belief we could try describing their underlying physical processes and avoid referring to their content, or to the purely abstract concepts with which they seem to be entangled. We would, if such a procedure were possible, have thereby dispelled or bracketed the abstract parts of belief and desire, the parts represented by truth, meaning, and value. But this procedure dissolves the very idea of a belief or a desire. Further, it assumes, as philosophers often do, that truth, meaning and value are purely abstract. I suspect that the assumption that leads us to think of these and other basic concepts purely as abstractions stems from an historical desire of philosophers to give an account of every concept they touch that will stand a chance to achieve final clarity. Such a desire, or unacknowledged methodological aim, is futile if the concepts in question are essentially interrelated *and* if they (e.g., truth and meaning) are not merely abstractions but are wed to the underlying intentional reality of language and thought, in other words, to use a partly redundant expression, to a linguistic-*intentional* consciousness.

Let us focus for a moment on meaning. Most philosophers of language and linguists would agree that meaning cannot be understood in isolation, apart from a language. But there have been a variety of different accounts of language, including those that ask us to focus almost exclusively on a language's structural features, and to conceive these in abstract terms. A structurally oriented account of language might be useful for certain purposes. But if such an approach leads us to regard the meaning of a language as abstract, we will miss an essential purpose of language, which is to communicate meaningfully about the world or our experience of the world. This purpose links meaning to the intentional concepts we're discussing. For meaning is never communicated until it is understood by an actual person, perhaps reflexively by the speaker, and in *the act of understanding* it becomes something more than an abstraction; it becomes a series of intentional events, of interrelated thoughts.

Meaning cannot exist in isolation from language, and a natural language cannot exist apart from the users (speakers/thinkers) who understand the world with it, or with their version of it. Conceiving meaning in isolation, either from a language or from the speakers who speak and understand with it, might be useful for the limited aims of a purely analytic or linguistic inquiry, but an ontology of language needs to coordinate an account of language, speaker/thinker, and the world *together*. Since an account of meaning is subsumed by an account of the language, the ontology of language and meaning presupposes an account of their actual existence in the consciousness of the speakers/thinkers who communicate their perceptions, sensibility and view of the world. This ontological implication suggests why we should not regard meaning merely as an abstraction, and represents one side of our proposal that the concept of meaning is irreducibly complex.

The complexity of meaning, grounded in the changing worldviews and consciousness of language users, makes the interrelated concepts of truth, value, belief and desire more conceptually dense; and in turn meaning becomes conceptually denser as a result of these inter-relationships. These concepts can only be understood concretely in relation to the worldviews and *consciousness* of language users, and therefore only as part of a constellation of interrelated concepts that are each irreducibly complex. The other side of our proposal, which I won't begin to develop in our discussion, is that consciousness is essentially and exclusively a feature of animate, biological beings, which further complicates intentionality.

3. Free will, language and radical interpretation

We should think of free will as part of this constellation of interrelated and embodied concepts presupposed by thought, at least by higher thought; for without it many of the novel acts of understanding of which competent users of a language are capable would be impossible. This premise becomes more plausible when we look at the cognitive-linguistic activity of radical interpretation. *Radical interpretation* refers to communication between individuals that overcomes the obstacle of basic (radical) divergences in their respective worldviews and corresponding linguistic commitments, where their understanding of language and the world has them viewing and making decisions in the world from incommensurate standpoints. The idea of radical interpretation assumes that there are meaningful utterances or expressions that neither a perfect understanding of the semantic conventions of a shared language, nor perfect facility with the ordinary uses – or even all pre-existing shared uses – of these conventions in a linguistic community can capture. Such utterances would presumably count as novel from the standpoint of the interpreter, until or if she captures their meaning and situates it in the context of her own language, which would need to expand in the process.

There are many instances of comprehension of novel utterances that would not count as instances of radical interpretation, for example when a speaker introduces a clearly specified new term and then uses it in a sentence whose context is evident, or when she uses language to bring to the attention of an interlocutor some novel feature of the world without the aid of any shared terms or formulas, conventions of any kind that are unfamiliar to the interlocutor understanding the utterances or calculations that these terms, formulas, etc. rely on. Understanding even novel utterances in such scenarios perhaps does not show conspicuous agency, at least not if all the terms and relations between terms can be analysed or explained without adjustments in the interpreter's current understanding of the world and language. I take it that such an understanding could be mimicked successfully by an interactive algorithm. But if significant cognitive-linguistic adjustments are needed for communication and understanding to occur, then the need for choice would seem to follow. The word 'choice' is perhaps ambiguous; these days it might refer to an automated selection procedure, as opposed

to the conscious choice of a person. Automated procedures obviously are unfree. But even conscious choices are not necessarily free, since a conscious creature or an agent might be only vaguely aware of a selection that it makes, or operating under an irresistible compulsion. Free will, if it exists, arises instead with choices that are inseparable from consciousness, in which the choice is deliberately made. That at least would seem to be a minimal condition for free will.

The deliberations of a radical interpretation who is trying to understand a novel point of view suggest an existential possibility. The radical interpreter might change herself in the sense that she might change her worldview, and therefore her self-identity. A lasting or temporary existential change of this kind occurs when an interpreter, confronted with alien linguistic behaviour (e.g., an utterance or expression), is prompted to revise her linguistic commitments, and underlying beliefs and values, with which she understands the world. Success in understanding an alien meaning thus brings about a new relationship between herself and the world. A supposition of agency is only required if this change was brought about by her making a conscious decision. We might consider the possibility of a selection procedure or 'choice' in which existential-cognitive changes similar to those incurred by the radical interpreter occur mindlessly. We can get a better sense of this possibility by examining two models of freedom that each differ from the kind of free will presupposed by radical interpretation: first, a universalist model, which depicts freedom as aiming at a single moral vision; and second, an utterly relativistic, idiosyncratic model, which severs the intentional dimension of one's freedom – what the beliefs, wishes, hopes, and so forth underlying one's choices are about – from a reality beyond oneself, which is irreconcilable with the primary purpose or radical interpretation: to better understand the relation between an alien language and the world.

4. Universality and freedom

Radical interpretation, based on the interpreter's unique beliefs, values, proclivities, sensibility, worldview and language, and involving these in a multitude of interpretive decisions that she makes while trying to understand an alien meaning, presupposes free will only if her interpretive decisions are (i) self-reflective, (ii) inventive or creative, and (iii) governed by reason, minimally by a consistent understanding of reality.

Kant's universalist model of freedom emphasises this last constraint, famously in his first formulation of the categorical imperative, which offers a selection procedure to keep a person's moral choices safe from the contaminating influence of blind impulses and prejudices, by bringing these choices in line with a set of consistent, universally binding and applicable moral principles. This model seems to be irreconcilable with the idea of *individual* freedom. Consider an interactive algorithm designed to derive moral principles from a categorically true starting principle – e.g., 'You should help persons where time and consistency allow and respect their

(negative) rights in so far as they respect the rights of others.'[5] The algorithm, let us stipulate, can populate a set of moral principles based on this starting principle more efficiently than any person. It would be reasonable to allow that the resulting set of principles it logically derives represents a perfectly harmonious array of moral truths, which, in line with Kant's universalist aims, provides a basis for eliminating all idiosyncratic beliefs which can't be brought into logical alignment. A person who was made aware of this wondrous algorithm might choose to accept its completed set of principles, at least if she believed that the starting principle was categorically true. But even if she had every reason to believe that the successive principles of the set were flawlessly derived, her choice would hardly exemplify our normal idea of freedom; for she would not have chosen these other principles but only the principle from which they were derived. What if her decision was self-reflective (it satisfies 'i'), and she reflected critically on her acceptance not only of the starting principle but of each derived principle, to assure herself that she, not the algorithm, was choosing the resulting core beliefs of her moral worldview? She might, intuiting that free or deeply free choices are thought to involve inventiveness and creativity (ii), examine each principle *imaginatively* in relation to clumps of beliefs and values that define her unique moral outlook. If she satisfied 'i' and 'ii,' her decision to accept these principles might start to look more like a free choice. But if she is committed to reason, minimally to consistency (iii), she might wonder at the purpose of this elaborate task, and ask herself whether she could spend her time more productively on other cognitive-existential tasks or problems. If she is rational, accepts the starting moral principle as categorically true, and understands that the moral principles she is considering are flawlessly derived, she would already have a compelling reason to accept all the principles in the set of principles that the algorithm presents. Beyond detecting and eliminating, or re-imagining the role of, idiosyncratic beliefs and values which she happens to hold and which aren't part of the algorithm's set, imaginative (ii) exercises in self-reflection (i) would seem to amount to a needless and counterproductive expenditure of her (morally) precious time and energy. From a moral point of view, she might better exert herself simply activating the moral program, and encouraging or requiring herself – perhaps permitting herself to be forced if she's too weak willed – to eliminate summarily her old beliefs and begin implementing the adjustments in kinds of behaviour required by the (possibly expanding) set of universal moral principles that her devotion to reason requires her to accept. Free will, if it leads her towards expenditures of self-reflection beyond her initial reflection that the selecting algorithm she has accepted begins with a universal principle and is consistent, would be wastefully inefficient; and it would be ethically misconceived if it draws her into labyrinths of imaginative reflection over idiosyncratic beliefs and values that interfere with her moral obligation to align her character harmoniously with the results of the algorithm. Instead of serving as an ongoing requirement of moral life, her freedom of the will in the sense suggested by our minimal conditions would become an impediment.

Should we then adopt a universalist conception of freedom? We might accept that an individual's rigorous incorporation of universal moral precepts into her moral thinking and character has released her from the grip of inclinations stemming from biological impulse and social prejudice. There is a sense in which blindly following inclinations can enslave a person, which might lead us to accept that her successful efforts in releasing herself from the control of blind impulses and prejudices could make her more autonomous. None the less it seems strange to say that she is less free if she follows her inclinations. A Kantian program need not rule out her inclinations, which would make it psychologically implausible. She may regard her inclinations as morally acceptable if she has reason to believe (perhaps she checks the algorithm) that they are categorically conforming to a set of universal rules, or at least don't violate any universal rules. But this qualification overlooks an apparently important aspect of free will, which is that it involves the more or less ungoverned play of inclination. Without assuming this aspect of cognitive-emotional play, can we adequately conceive of free thought and free action? Our Kantian rule follower might have *chosen* the rules that she is duty bound to follow, in the sense that she is assured that all successive rules of the algorithm she follows are consistent. But it would still be strange to say that her later choices are free if, after her initial choice, she merely follows these rules. If she follows these rules categorically, her successive 'choices' would, though they are rationally well-grounded, appear to be more automated than free. Indeed neither she nor we would be able to distinguish her choices from those of an automated program designed to 'choose' on the basis of the same set of rules.

Whether or not the concepts presupposed by thought are essentially interrelated, reason is more involved activity than Kant's first formulation of the categorical imperative suggests. As a form of (conscious) thought, reason must be embodied and hence integrally tied to inclinations and passions, which can't be disentangled from belief, value, meaning, and so forth. A person animated by reason in this conceptually and physically involved sense might freely decide to govern her existence by a set of universal rules. But that she chooses a succession of universal principles because she knows they are consistent with this set does not begin to show that she has chosen freely. These choices would only resemble *free* choices in so far as they imaginatively involve her individual sensibility and own way of viewing the world, on the basis of the intentional dimension that she uniquely experiences through her feelings, intuitions and beliefs. A principle such as the first formulation of the categorical imperative, applied universally to every rational being, obscures the complex and messy nature of the intentional activities in which free thought emerges. It is incompatible with the idea of a *single individual* – who becomes redundant if there are many exemplary rule followers, living morally identical or similar lives. Not incidentally, it is also incompatible with the view that a free individual will tend to stand apart from the crowd, the likelihood that an individual whose freedom is highly realised will tend to diverge from many widely shared moral and other views, overtly or within her own thoughts, no matter how exemplary her peers might be.

There are universal moral principles and imperatives that, whether recognised or not, underlie every rational individual's worldview (e.g., 'Truth is a value' and 'Aim to avoid inconsistency'), and others that have come to be widely accepted within certain civilisations (e.g., 'Government should protect the liberty of its citizens' and 'Treat strangers with respect'). Though some of these principles might be entailed by the categorical imperative, by themselves they do not explain very far the thoughts and deeds of individuals who exemplify freedom. Programmed into a machine that has been designed to produce morally flawless 'decisions' in certain well-specified contexts, these principles might let us understand and predict the machine's output perfectly (perhaps with the aid of another machine). Within the thought of a person, no principle, belief or value could function like that, for a basic linguistic, ontological reason: A principle, belief or value mirrors the role that a sentence plays in a language, a language that represents the (ever-changing) linguistic commitments and thought of an actual language user. Just as we cannot make much sense of individual words, linguistic fragments, or sentences apart from a language, an individual's professed principles, beliefs and values signify nothing, or very little, in isolation from her worldview. When we refer to *significance*, already we are assuming a language, which is inseparable from the thoughts and worldview of individual language users. When we speak of either language or of thought, the other is pervasively being assumed. And once we take into account the dynamic and holistic nature of thought or language, it becomes hard to maintain the view that individual autonomy is underpinned by an individual's categorical reliance on a set of universal moral rules which can be isolated from the entirety of her thoughts.

We see the limited use of moral rules to explicate the idea of freedom when we consider what it means to think and act freely on the basis of Kant's second formulation of the categorical imperative, the moral rule that *we should categorically respect persons*. Following a categorical rule that she has given to herself is meant to release a person from inclinations which tie her to the deterministic order, which might cause her to act contrary to reason. The only part of this proposition that relates interestingly to freedom is that she gave herself the rule and that it is informed by her view of things. A person or a machine might act in accordance with the universal principle 'We should respect persons' through habituated or automated thought, or be programmed to accept the principle, and keep it perfectly by refraining from activity that violates any person's (negative) rights. The principle itself is inert. We can only relate the principle to the freedom of the individual who wishes to act in accordance with it by assuming that it operates in interaction with myriad beliefs and values and other wishes within her thought, and that she chose it because of a decision based on such an interaction.

Within her thought the universality of the principle dissolves in its application insofar as she is thinking freely. As soon as its application approximates concrete, integrated thought, at

thinking that isn't automated or routine but fully integrated into her unique configuration of beliefs, values and ways of thinking, its universality slips away. In itself the principle, in addition to being inert, is an abstraction, a semantic shade isolated from the meaning of any actual individual's worldview. When an individual applies the principle as an abstraction, without fleshing it out in her own thoughts and in relation to the particular (unique) person whom she intends to respect, she is hardly thinking, let alone thinking freely. If instead her thinking represents a high degree of autonomy, it will have turned the universal principle that she is applying into a thought whose meaning is progressively more particular to her beliefs, values, sensibility, and outlook. This is one side of the hermeneutic act required if the abstract principle 'You should respect persons' is to refer to actual persons. The individual applying the principle also needs to coordinate her thoughts with (a projection of) the intentional reality of the person whom she decides to respect. Her thinking cannot remain abstract if she is thinking very far about another *person*, and not merely projecting thoughts that capture any person at whom they are aimed. There is no general person whom her thoughts of respect could capture if they only satisfy generally applicable criteria of identity. For there is no universal object of her intent; there is only another actual individual, who is not contained by the general idea of a person, and who therefore is missed by any universal principle, unless that principle is expanded into a voluminous particular statement, which would outstrip its original content.

5. Radical interpretation and self-recognition

It may be more difficult to respect an actual person than many of us imagine. I sometimes meet others superficially, and scarcely *perceive* them as persons. If my view of others were confined to behavioural criteria, I might begin to think that the persons I meet superficially, or over a lifetime, are reducible to bodies run by a natural algorithm, sustained by the neural networks that operate in us all. That is also how they would perceive me if they were to become consistent behaviourists. Yet, even in our most superficial encounters, this is not how we normally regard each other. Intuitively, without inferring or explicitly knowing the invisible phenomenology of our respective souls, most of us immediately recognise one another as a person.[6] Radical interpretation typically comes to the fore in moments when another person begins to speak or act in ways that stymie our immediate recognition, and when we become curious about what the other person said or did or how they behaved, which has struck us as baffling, mysterious, grotesque, incoherent, meaningless, and so forth. Such moments are inevitable if we are ever to know another person very far, since the worldview that underlies a *person* is unique and unfixed, and perhaps permanently elusive. But can we know or recognise others, or ourselves?

The belief that we can recognise another person is epistemically ambitious; it assumes more than our behavioural evidence allows. If we lacked the assurance of our own conscious

experience, we might despair that we ever do perceive persons, or any conscious creatures. Abandoning our epistemic assurance that we are conscious has become one way to deal with the problem of recognition. We simply consign the concepts of consciousness and recognition to the flames. But trying to maintain the view that I am not conscious is difficult, and immensely impractical, which might be why some philosophers welcome Dennett's proposal that we need to maintain a systematic epistemic policy that our intentional concepts are to be regarded as true of reality, even though they really aren't. Dennett's strategy, it seems to me, crosses the line into sophistry, and the problem it is meant to address is not merely impractical. If I could convince myself that I am not conscious, I would be left with a profound epistemic problem, which of course I would be incapable of understanding. I can only justifiably convince myself of the sceptical view that I am incapable of conscious thought if this view is false; and I would be in no position to convince myself of this view, or of any problem or proposition, if it were true. From my personal viewpoint, I should regard the scrupulous behaviourism that asks me to abandon the assumption that I and others are conscious as absurd.

The problem isn't confined to my personal viewpoint. For while my brain and the brains of others might *unconsciously* produce and share a system of symbols that represent and coordinate the stimuli that activate our respective organs, none of us would be capable of attaching sense to the symbols if behaviourism were true. Let us allow that these symbols, when received as differentiated stimuli by the sensors connected to unconscious brains, cause the assemblages of neurons that support the bodies of which they are a part to perform their operations. These brains could then *interactively* 'process' huge sums of data which would let them coordinate and communicate to others a 'view' of the exterior world, which might stimulate them to behave in wondrously familiar ways. Though this data would be senseless, in theory it might still produce behavior in a community of these unconscious beings indistinguishable from the behaviour of conscious, thinking persons. Their community would be ripe for an elaborate Turing test, although there would be no one to test if consciousness is an illusion. If we are nothing more than these unconscious brains, would we be able to speculate whether our sensory-neural operations can generate and exchange the kind of sentences that we are processing now, which refer extensively to an intentional reality? If the foregoing thought experiment described the only possible reality available, these sentences would refer to quite a few illusory concepts. They might still be exchanged in a manner of speaking, but there would be no exchange of actual concepts if concepts are inseparable from thought.

*

Since my thought, if it extends as far as beliefs and views, is seamlessly part of a language, my idea of myself assumes the idea of others. For language, as nearly everyone

agrees, is an intersubjective activity. Even if none of us speaks and thinks entirely or always with the same shared language, we would be unable to speak or to think in the unique ways that each one of us does had we not developed in a linguistic community; nor, after we acquired a language in this community, would we be able to continue with the cognitive life this language enables if we did not assume some actual or hypothetical other as we formulate our thoughts and points of view. This other might only represent an aspect of ourselves; but our intentional landscape, language and worldview are so thoroughly saturated with the beliefs, values, and ways of perceiving, feeling, and thinking of others that this is a moot point. Regardless of whether or not we address some other real or imagined person in particular instances, every particle of our thought continuously relates us to a wider intentional world of thought and feeling, which we largely inherit through our interaction with others and the utterances, behaviour and artefacts they have created or embody. Our linguistic, cognitive, and perceptual inheritance is as extensive as we are; and if we develop or change, we do so amidst the direct or transfigured influences of others, even though the unique aspects of our personal worldview and sensibility are what differentiate us and define who we are as individuals.

That our identity as self-conscious beings is both communal and individual suggests a tension in our idea of the self. The influence of the language and worldview of others on the self is as pervasive as our biological inheritance; it permeates every aspect of our soul, and can't be entirely assimilated or transformed. If true, this claim narrows the prospects for agency, and hence for the very idea of the self. Agency itself is one of the foundational existential categories and problems of our tradition, in which the self has been conceived as a self-conscious being capable of free deliberation and choice, and thus potentially responsible for her behaviour. Faced with this conception of the self, we might easily come to regard the immense influence of history and language, the cumulative depth and flux of other worldviews that these connected phenomena represent, as a threat to the very idea of the self. For by all appearances the capacities that allegedly allow the self to deliberate freely have been determined by the diverse phenomena of history, language, and a great deal else. If the idea of the self presupposes the idea of agency, it might seem that both ideas are in jeopardy.

But should we think of the self as threatened by the conditions from which it allegedly emerges? What conditions would permit the self to emerge safely? Perhaps we should try to imagine a self that has created the grounds of its own existence, or which creates its being from nothingness. In a discussion that aims to move beyond the traditional problem of free will, Nietzsche dismisses "the desire for 'freedom of the will' in the superlative metaphysical sense" as involving "nothing less than to be precisely this *causa sui* and, with more than Munchhausen's audacity, to pull oneself up into existence by the hair, out of the swamp of nothingness" (BGE, 21), a desire which, he says, "still holds sway . . . in the minds of the half-educated."[7] I doubt that anyone actually ever maintained such a desire, or could have.

Nietzsche intends this characterisation of free will to serve the rhetorical purpose of clearing the conceptual ground. His primary intent is to attack the contrary idea of an "unfree will," which, in his view, "amounts to a misuse of cause and effect" by those who seek to "naturalise in [their] thinking": such thinkers "wrongly reify 'cause' and 'effect'" by making the cause press and push until it 'effects' its end."[8] Nietzsche, unfortunately in my view, councils that we "should use 'cause' and 'effect' only . . . as [theoretical] fictions for the purpose of designation and communication – *not* for explanation."[9]

Less sweepingly, we might leave the issue of standard causation alone and ask whether all the decisions that an individual makes are necessitated by circumstances outside herself. The answer should be obvious in one sense. A person is entirely in the world and never would have come into existence without the total circumstances of the world. These circumstances necessitated every particle of her existence, including all her decisions, past and ongoing. All this is true if we regard persons as part of the natural world, or if we have no reason to believe in a metaphysical distinction between subjects and objects. But persons exist temporally, as a process, and change over time; and over time, the nature and locus of the circumstances that shape their identity also change. As an individual's identity takes shape, her capacity for autonomous decisions develops. An emerging embryo exercises no autonomy. If it grows into a normal adult, the person who has developed from that embryonic entity will make decisions that are autonomous, to some degree or other. Nothing in this familiar account interferes with the claim that no person creates the circumstances of the world that bring her into existence, which simply affirms the point that no degree of autonomy lifts her out of the totality of all being. If we accept this point, we can consistently say that some of her decisions are born of conditions that include her own self-conscious thoughts. The world gives an individual the capacity to think autonomously, and in her development, by degrees, she exercises this capacity; she also contributes to its development, with every decision she makes that affects her core beliefs and values. Decisions such as these change her identity, but nothing about her temporal-intentional condition, of which these decisions are a part, implies that she is thereby transcending the world by degrees.

She remains part of the world, but not as an inert ghost passively watching the causal activities that stir and create the underlying phenomena of her person. Over her personal history, an individual becomes free to act in ways that play a role in shaping who she becomes; by thinking and acting in certain ways, singularly and through cumulative choices, she increases or diminishes her capacity to act freely. The circumstances that bring her decisions about include her own intentional skills, a capacity to discern and to become aware, to interpret and understand. The world that bequeathed to her these capacities necessitated the decisions by which she exercises her freedom. And since she is a uniquely influential part of this world as it affects her, she has helped bequeath these capacities to herself. Without forgetting the fact

that she is part of the world – which would be hard to do in such a discussion – it seems natural to say that she has helped create some part of herself and thereby controls some part of her fate. Speaking this way makes it natural to say that, precisely on account of the necessitating circumstances of the world, she sometimes decides freely and in some measure is responsible for her actions.

*

The freedom of any person is extremely limited, extending only as far the skills, proclivities, character, and labyrinthine content of beliefs and values underlying the intentional conditions that bring about, or undercut, self-conscious decisions. I say above that unique conditions within a person help bring about her current appearance in the world, and who she has become. But the world is full of phenomena that possess unique features and participate in their own birth or change – from interactive neural networks and differentiating cells to wave patterns and weather systems. Though lacking the intentional conditions tied to consciousness, a mindless device or process might bring unique aspects of itself into the world of which it is a part. But without an intentional attitude, a point of view, no thing or process can be autonomous.

Perhaps no individual is born with a point of view. Certainly no individual springs into the world reflecting on a worldview that underwrites her capacity for autonomy. An individual's capacity for autonomous thought and deeds depends on a protracted personal history, which she only gradually and precariously achieves. As I suggest above, the worldviews of others is an essential part of this intentional history; and their perspectives potentially bolster or retard its progress. A person's intentional history is uniquely a product of self-reflection and self-recognition. But the intentional history in which self-recognition appears depends on alien points of view that an individual has absorbed, either by incorporating these perspectives relatively unchanged or by transforming them into unique pieces of her identity. Instead of *worldviews* and *points of view*, we could refer to *beliefs*, *values*, *sensibilities*, *ideas*, and other intentional concepts. We might be tempted to refer to *ideas* in another sense of the word, or to *propositions*, *content*, *data*, *information*, *storag*e, etc. But with these substitutions we would obscure the *directed and subjective* aspect of the interrelated elements of an individual's worldview, which give life, literally in some mysterious sense and also in a figurative sense, to entities that otherwise are abstractions. Because the beliefs and values that an individual absorbs into her private history from a wider history are inherently intentional, and as such represent *perspectives*, they accumulate within her forms of recognition that in some sense remain other than herself, and not merely as a *content* that she disinterestedly accepts or rejects. Assuming that her beliefs and values never congeal into a categorically harmonious whole, contrary perspectives will abound in her worldview. These opposing forms of

recognition, transmuted or approximately preserved, will have become part of her identity, influencing and complicating her outlook and decisions.

6. Mixed selves and self-alienation

In section 260 of *Beyond Good and Evil*, Nietzsche announced that among the wide variety of moral perspectives "which have so far been prevalent on earth" he had "discovered two basic types": "*master morality* and *slave morality*."[10] He "immediately" added this complicating qualification:

> [I]n all higher and more mixed cultures there also appear attempts at mediation between these two moralities, and yet more often the interpenetration and mutual misunderstanding of both, and at times they occur directly alongside each other – often in the same human being, within a *single soul*.[11]

If we conceive of persons as individuals who have developed their inherent freedom to the extent that self-reflective moral thought has become second nature, we should expect to find within the moral worldview of a *person* an "interpenetration" of a variety of incommensurate moral perspectives – or contrary forms of moral recognition "within a single soul." If we actually could peer inside others, we would of course also find an array of extra-moral – e.g. metaphysical and aesthetic – perspectives merging and competing as part of the intentional conditions that create their thought. An individual without such a mixture of contrarily directed perspectives within her soul would be a comparatively impoverished *person*, assuming the term could still be meaningfully applied; for the intentional conditions that underlie the degree of freedom of thought of an autonomous being would be absent. Absent would be the creative mixing of perspectives that enable the individual to renew or overcome her current values and outlook; also absent would be the means by which she might discern the incompleteness or relative poverty of her soul, which provides a sufficient motivation for a highly autonomous person to want to change.

While the contrariety of perspectives that pervades "higher and more mixed cultures" lays the groundwork for an expanded freedom of thought, it is not obvious that an individual harbouring myriad irreconcilable perspectives will exemplify freedom. Such an individual, if burdened and self-alienated by inner contradictions, might as readily succumb to a paralysis of thought that suspends the freedom of her soul. Her attitude toward particular incommensurate views which she experiences as alienating or burdensome would matter – e.g. how tolerant she is of dissonance, or of indeterminacies or an unsettledness among her beliefs; or how spiritually cautious or venturesome she is, or lazy or industrious, cynical or hopeful, lifeless or passionate, fearful or courageous, apathetic or curious, closed or open, and so on. To think and will freely, she would need to possess, along with the critical interpretive skills

required to unearth and understand the meaning of myriad points of view that might prompt her to examine herself, a willingness to remain open to the likelihood or reality of self-alienation. This potentially free person's explicit, sustained moments of self-conscious freedom require in other words the open disposition and reflexively critical skills of an internally directed version of radical interpretation.

7. Charity at home

This open disposition is formally encapsulated by *the principle of charity*, which is the methodological basis of radical interpretation. The principle of charity represents the theoretical assumption that an incommensurate language which an interpreter has begun to understand enables its speakers to understand the world. The principle prescribes that this interpreter attribute truth as far as possible to the language she recovers, as a precondition for understanding what sense it exhibits, especially when confronted by off-putting initial appearances, e.g., of obscurity, grotesqueness, confusion, nonsense, or obvious error. The principle is a continuous process, since the interpreter's theory of the language is never complete, nor entirely true. As she observes instances of anomalous or strange linguistic behaviour, the principle provides the theoretical impetus for her to reconfigure the pattern of meanings with which her theory aims to capture the language, and to adjust linguistic commitments and beliefs in her own language and worldview insofar as they prevent her from constructing a more plausible theory. The only constant in this continual reconfiguration of the interpreter's commitments and of the presumed language she recovers is the assumption that it is true of the world – true in the sense that its salient concepts refer to the world, and in the sense that it allows those who speak it to make true and false assertions about the world. The principle takes truth as an absolute, and the incommensurate languages-worldviews that an interpreter begins to coordinate as provisional approximations of each other, and of the world, in need of charitable revision.

A person who consciously overcomes aspects of her own worldview will have recourse to a self-directed version of the principle of charity when she examines and tries to understand incommensurate views within herself. For the sake of contrast, let us assume that she has passively received all her beliefs and values from her cultural environment, without a vestige of critical awareness. From this unpromising starting point, can she overcome her worldview and cultivate a new self? If her new self is not essentially a reiteration of her former self, nor an expansion that keeps in place the elements of an inheritance for which she isn't responsible, her act of self-overcoming will need to develop tensions among her current configuration of beliefs and values; and she must consciously influence the formation of her new configuration. Perhaps these tensions will dissolve naturally, without conscious effort, while she undergoes changes. Self-overcoming, I think we should concede, largely does occur as a result of mindless

processes occurring within oneself, e.g. in one's central nervous system or among beliefs from which an unbidden idea or inspiration emerges. If unconscious activities tell the whole story of self-overcoming, however, the self will have changed like a weather pattern might change, and the individual who has changed will be no more responsible for her new incarnation than she was for the old. For her to become responsible for the intentional conditions on the basis of which she decides and acts in the world, she would need consciously to have chosen some aspects of her worldview, to have consciously transcended some part of herself; and to do so she will need to have created or unearthed tensions in existing aspects of her (inherited) worldview, and enacted a process of self-examination similar to the process prescribed by the principle of charity.

Need an individual remain *unfree* in the sense that implies responsibility if she suppresses or ignores, or is oblivious of, tensions within her that might, if attended to, precipitate change? By not attending to such tensions, she will have narrowed or left undeveloped a range of perspectives that might have presented her with a wider array of choices, which would seem to leave her comparatively less free. But that concern is premature if she is not yet free in the required sense. A more basic concern is that she remains uncritically attached to perspectives that were determined by external factors, or by mindless processes within her for which isn't responsible. She will stay unfree until she acquires the critical perspective needed to assess, and choose, core aspects of her worldview. Selecting beliefs from an entirely harmonious range of perspectives within her would keep her unfree if they belong to a worldview she didn't create. Might she release herself from aspects of this worldview with a spirit of contrariness, or a fertile imagination? The question supposes that her worldview isn't harmonious after all, unless her contrariness or imagination are content free. But of course no worldview or self is entirely harmonious. Selves and their worldviews are inherently sources of tension – bundles of contradictions, as Hume says. An individual who, by design or neglect, *prevents* tensions from arising among her beliefs, or fails to consider the significance of the tensions she perceives, might benefit from a spirt of contrariness, coupled with an imagination. But the principle of charity retains an essential role to play in self-overcoming if the individual is to become responsible for the beliefs that guide her thinking, at least if charity is conceived as *a precondition of understanding*.

A sceptic of free will might insist that, however resourceful the individual's procedure for selecting the elements of her worldview becomes, determinism is universal and so any apparent expansion of her freedom through self-overcoming will be based on an illusion – the illusion that she can release herself from the antecedent circumstances that created every aspect of her, including all the beliefs on the basis of which she acquires further beliefs. In section 5, I sought to weaken this common argument by observing that *she* might be included among these circumstances. I will add now that the determinist argument against free will

equivocates over circumstances that determine the content of the individual's beliefs and those that also determine her array of higher cognitive skills or skills in reasoning. These skills enable her to assess, from a relatively disinterested standpoint, the beliefs she acquires, including methodological beliefs which can enhance or impair the operation of these skills. Whether or not she created the circumstances that bequeathed to her these skills is irrelevant to the issue of whether they enable her to think and choose independently.

This point can be extended beyond the issue of skills and informing methodological beliefs, to the issue of non-methodological beliefs. It should be extended, as her reasoning skills cannot work in isolation from the rest of her beliefs; they depend on the totality of her beliefs, which might impair the function of these skills, or help facilitate their proper work. The dependency of reasoning skills on belief isn't cause for despair. For any belief that she has acquired, regardless of its genealogy, might foster independent thought, e.g., by countering the influence of other inherited beliefs or prejudices. That beliefs are part of a deterministic universe does not diminish the role they might play in support of free will. Free will *depends* on the deterministic efficacy of beliefs and values, and related dispositions. The issue is whether these remarkable intentional dispositions can combine in a variety of ways with other circumstances in the world to produce an agent capable of influencing her own intentional states, without simply perpetuating the existing beliefs of her inheritance. The issue seems to come to a head with this question: Is she or isn't she locked into the worldview in which she finds herself? The answer isn't obscure, certainly not if we have already relinquished any attachment to the subject-object distinction. She can't be locked into her worldview if she is able to understand others who view the world differently, and if she wishes to understand their differing viewpoints. With this ability and attitude, if in other words she is animated by the principle of charity, she might expand her worldview indefinitely, and expand and exercise her freedom in the process.

8. Apparent reversals

The capacity of human beings self-consciously to change and expand the conditions of their freedom appears to be a unique accomplishment of our species. The chorus of Sophocles' *Antigone* might have begun by extolling this accomplishment when it elaborated the *wonders* of man. For the most *wondrous* (awesome, uncanny, terrible, etc.) of his accomplishments relies on the nature and quality of his freedom, or to say much the same thing, on the nature and quality of his thought.

If thinking and freedom are inseparable, we should hesitate to dismiss the possibility that members of species that are incapable of deliberately affecting the conditions of their being are therefore incapable of experiencing freedom. If there is no way to separate the interrelated intentional elements of thought, then we should allow that non-deliberating

animals, humans included, are capable of a freedom extending in degrees, as far as their ability to think. Might a machine also be capable of freedom and thought? That would depend on the nature of the machine. We should affirm this strange question if the machine consciously *embodies*, and doesn't merely imitate, belief, desire, meaning, and the rest of the interrelated conceptual-emotional phenomena that thinking presupposes – if it has become biological and animate. But only with self-conscious, deliberative thought does a higher freedom appear on stage, the kind that carries with it the blessings and pitfalls of moral responsibility.

Self-conscious thought, though, might hamper freedom while it expands its basis. A creature such as an ass, a bat, or a lion, or a conscious (biological) machine such as we have yet only encountered in works of fiction, might, without deliberative, self-conscious freedom, possess intuitive or computational powers that exceed in fluency, complexity and precision the thought of a self-reflective person in many situations. These conscious creatures, or a singular machine, might experience their freedom more naturally – perhaps I should say *with greater facility or fluency or ease*. Consider the situation of an individual who descends into reveries of self-reflection in which she begins to change prominent elements of her worldview. She is liable to incur far-reaching inefficiencies of thought and action as she changes, especially if her reflections remain dissonant, or until her new outlook has become second nature. Since she would need to cultivate this situation in her life more or less continuously to sustain the highest states of freedom, she might, as Nietzsche appreciated,[12] wish instead to consolidate or limit the intentional conditions that increase her freedom, or seek to preserve an equilibrium among her perspectives and desires. A tradition might facilitate this constraining wish if its inherited resources help soften the effect of newer outlooks, or its imperatives discourage her from venturing beyond her existing horizons.[13] The intersecting traditions of "higher or more mixed cultures" are more liable to encourage the persons whom they engender to change their worldviews, and enable them to recover their equilibrium more readily when they undertake such change. This cultural eventuality suggests why we should reject the deterministic *conceptual* inference that the causes of our creation must subvert our freedom.

We should also reject the endpoint of neo-positivism or empiricism, which encompasses the dogma that thought itself is an illusion, or a phenomena that can in principle be reduced to simpler phenomena. This dogma offers a paradoxical position, as it assumes the efficacy of a faculty which it dissolves. Many intrepid neuroscientists, neuro-philosophers, cognitive scientists, and psychologists will unlikely be deterred by this paradox from arranging experimental inquiries designed to underscore what they already regard as the illusory nature of *thought*, *free will*, *consciousness*, and a host of related intentional concepts. All such inquiries will require intrepid investigators if these concepts are interrelated in the ways we have discussed, and if together these concepts underlie every *point of view* we care to take in any inquiry. The positivist belief that the real-world correlates of these interrelated concepts

are inseparable from the goings on of our bodies might seem to hold a distant hope for the confirmation of their doctrine that the phenomena of thought are reducible to things "simpler and more fundamental."[14] Their belief that there are no disembodied thoughts is true, by all the evidence that any human being is ever liable to secure. But that this belief should lead them to the doctrine that belief, desire, hope, intention, agency, and all the other phenomena that underlie thought are *reducible* to mindless physical phenomena is a testament to the perennial poverty of a positivist's logical imagination.

---

[1] Elsa Youngsteadt, as quoted by Alfred R. Mele in his debunking book *Free: Why Science Hasn't Disproved Free Will* (New York: Oxford University Press, 2014), page 26.

[2] Dennett explicates this view with great care in "Real Patterns," The Journal of Philosophy, Volume 88, Issue 1 (Jan., 1991). In recent works, he describes the self and its intentional properties as "a theorist's fiction," e.g. in "The Self as the Center of Narrative Gravity," *Intuition Pumps and Other Tools for Thinking* (New York: Norton, 2013), or "Artifactual selves: a response to Lynne Rudder Baker," *Phenomenology and the Cognitive Sciences*, March 2014

[3] Davidson, "Truth Rehabilitated," *Rorty and His Critics* (Blackwell, 2000), ed. Robert B. Brandom, p. 73.

[4] *Ibid.*

[5] I take it that this principle captures the intent of Kant's second formulation of the categorical imperative.

[6] Roger Scruton fleshes out an impressive description of this sense of immediate recognition in Chapter 5 of his book *The Soul of the World* (New Jersey: Princeton University Press, 2014), pages 96-103.

[7] Nietzsche, Friedrich. *Beyond Good and Evil*, trans. Walter Kaufmann (New York: Vintage Books, 1966), Section 21, page 28.

[8] *Ibid*.

[9] *Ibid*.

[10] *Ibid.*, page 204.

[1111] *Ibid.*

[12] Nietzsche, Friedrich. *Untimely Meditations*, "On the Uses and Disadvantages of History for Life," trans. R. J. Hollingdale (New York: Cambridge University Press, 1997), Section 1, page 63.

[13] *Ibid.*

[14] Davidson, *Ibid.*

Works Cited

Dennett, Daniel. "Artifactual selves: a response to Lynne Rudder Baker," *Phenomenology and the Cognitive Sciences*, March 2014.

-----. "The Self as the Center of Narrative Gravity," *Intuition Pumps and Other Tools for Thinking* (New York: Norton, 2013).

-----. "Real Patterns," *The Journal of Philosophy*, Volume 88, Issue 1 (Jan., 1991).

Davidson, Donald. "Truth Rehabilitated," *Rorty and His Critics* (Malden, MA: Blackwell, 2000), ed. Robert B. Brandom.

Mele, Alfred, R. *Free: Why Science Hasn't Disproved Free Will* (New York: Oxford University Press, 2014).

Nietzsche, Friedrich. "On the Uses and Disadvantages of History for Life," *Untimely Meditations* (New York: Cambridge, 1997), trans. R. J. Hollingdale.

-----. *Beyond Good and Evil* (New York: Vintage, 1966), trans. Walter Kaufmann.

Scruton, Roger. *The Soul of the World* (New Jersey: Princeton University Press, 2014).