Persons, Virtual Persons, and Radical Interpretation
Michael Bourke

I – Personhood and the sciences

The concept of the person has been a concern of philosophy since philosophers began speculating over the Delphic maxim "know thyself." But it is becoming tempting to think that personhood and related concepts – e.g., intention, agency, value, belief, and mind – will not survive scrutiny from the physical and social sciences. Research in the overlapping fields of evolutionary psychology, neuroscience and cognitive science has perhaps already undercut the basis of the concept of the person, or shown it to be an illusion. That would be a remarkable event. For the concept of the person is not a concern of philosophers alone. The fate of this concept affects the wider field of human interests at a fundamental practical and conceptual level, underlying, as it does, our concepts of morality, our institutes supporting rights, justice, and the rule of law, our most intimate manifest image of ourselves, and the basis of the natural languages with which we encounter reality. When we consider the foundational character of this last interest, language, which underlies all our interests, it becomes clear, I hope to show, why we should reject as unintelligible the assertion that our concept of the person is illusory, notwithstanding what scientific research has to tell us about our evolutionary origins, genetic structure, neural and brain activity beneath our behaviour and thought, analogous computational processing, and so on.

I will start by considering a compatibilist theory of persons that suggests that scientific and technological developments only threaten an antiquated, implausible concept of the person, one which presupposes such burdensome metaphysical notions as an objective moral order and free will. From this standpoint, we might take comfort in the likelihood that research programs in these areas will someday coalesce around a revised conception of the self, based not on the illusory metaphysical presuppositions of linguistic traditions, in which we happen to find ourselves enmeshed, but on the harmonised results of approaches that will encourage us to re-evaluate the significance of our manifest, common sense self-image and to bring this image in line with "a scientific understanding" of reality. We might then, as that useful day arrives, have and keep our cake, use (revised) traditional terms associated with intentional concepts, such as value, personhood and agency, while eliminating their metaphysical associations. Daniel Dennett is a spellbindingly able advocate of this position.

The program of revising, in this way, concepts of our linguistic tradition(s) is tempting if the only alternative is to sit back and observe the incoming tides of unmediated scientific discovery as they erode basic moral, aesthetic, existential, and institutional commitments that depend on these concepts. In a recent book review, Daniel Dennett hints at the far-reaching significance of this problem when he says that the discovery that "free will is an illusion [would promise or threaten] to render obsolete a family of well-nigh sacred values: just deserts (for both praise and blame), guilt, punishment, honour, respect, trust, indeed the very meaning of life" (Dennett Oct 2014). Dennett clearly appreciates the practical stakes involved in dispelling central conceptual commitments of our traditional self-image, in this instance free will. With respect to free will and other foundational concepts, he has taken on a rather large public role as

mediator between our traditional and scientific images of the self. In this role, he has offered a basis for developing revised naturalised, non-metaphysical versions of central concepts of the manifest image of the self that our various traditions picture as metaphysical in character.

It might seem to be more sensible and clearer to leave aside the idea of linguistic *tradition*, and to refer simply to a tension between common sense and science, between our manifest and scientific images of reality. But there is a sense in which our common sense, manifest view has a *history*, which becomes embedded semantically, even syntactically, within our language. Those embedded linguistic changes that influence our common sense are what I have in mind by a tradition. Science clearly plays a significant role within our current tradition. But it may be premature to insist that a scientific point of view can, even in principle, exhaust our understanding of reality, or that we will ever be in a position to eliminate the metaphysical presuppositions of foundational concepts within our common sense tradition. Accordingly, in the discussion that follows, I will resist the rhetorical practice of Dennett and others of referring to concepts implied by ordinary linguistic use that do not align with a "scientific understanding" as the products of "folk psychology," as that convenient opposition encourages us to deal too offhandedly with the non-parochial metaphysical presuppositions of many of the basic concepts of our tradition, and of course to foreclose on the main issue under discussion.

II – The self as a "theoretical fiction" based on real-world patterns

Grammatically, writing about the self seems to incur a formidable challenge, since we can scarcely think or talk about the concept without referring to ourselves, and referring to ourselves as subjects. Further, the widespread view that language involves *intersubjective* communication is reinforced by this seemingly unavoidable presupposition: that the entities who use language, as a scheme for achieving mutual understanding, actually are subjects. If they were not subjects, it would be hard to imagine how their intentional acts of understanding, as opposed to the transmission of bits of data that *might* be understood, could ever come to pass. To avoid the implications of syntax or of language described as an intersubjective activity involving intentional states and agents, which force on us the existence of the thing and sorts of things we are questioning, it might be reasonable to regard the self as a hypothetical object, a device operating beneath considerations of ontology, until we can independently assess likely grounds for asserting or denying the existence of selves, or for revising our understanding of the concept.

Dennett's theory of personhood moves beyond this merely provisional stance, asserting that the self is permanently to be understood as a highly useful, theoretical construction, or, as he says, a "theorist's fiction" (Dennett March 2014). As a *theorist's* fiction, the self is not being put on a par with fictions which are mere inventions or illusions, e.g. Sherlock Holmes or unicorns. The kind of fiction Dennett has in mind instead is analogous to such theoretical constructs as gravitational centres or the Equator (Dennett 2013). By regarding the self as a theorist's fiction, he is neither eliminating the idea of the self nor adopting a realist stance toward selves, regarding them as unambiguously real. Instead he is advancing the position that

these two options present a dilemma which we can get past by accepting a doctrine of "mild realism" (Dennett 30-1, 1991). In effect we accept this view, when, using Dennett's sense of the term, we adopt "the intentional stance" towards actual (physical) or semi-actual (designed) entities; we commonly attribute intentional states to these entities without having an explicit opinion about whether these entities actually possess such states – beliefs, wishes, and so forth – or whether such states exist in some more substantive sense, physically or metaphysically. By formally accepting the intentional strategy, we can thus attribute intentional states to an indefinitely large range of entities – e.g., from hominids and bats to thermostats – and thereby cope somewhat successfully with their antics and behaviour, without worrying about the status of their underlying physics as we arrange our predictions, and without being seduced by murky metaphysical explanations. Instead, the intentional stance has us regard real-world patterns associated with such entities merely *as though* they were harbouring the presence of intentional properties. Since the stance lets us better predict the behaviour of these patterns and entities, unlike the creation of pure fictions, e.g. unicorns, an intentional concept can thus *sort of* refer to something real; it can refer to underlying real patterns onto which we only somewhat arbitrarily attach our theoretical predicates.

In taking the intentional stance we remain agnostic, perhaps mildly agnostic, about the reality of beliefs, wishes, values, etc., and the selves which collect these items in interesting ways. We only adopt the intentional stance because, Dennett says, the physical stance – which would have us trying to predict the behaviour of patterns that suggest selves from "neuroscientific levels of description" (Ibid 51) – would leave us with a computational headache. The intentional stance involves *treating* selves as real, even though, at the physical level – which trumps the intentional level (Ibid) and in principle eliminates it – they are not real, at least if we insist on speaking of everything as either real or unreal. In place of that antiquated doctrine, Dennett recommends that, in the special case of intentional objects, we accept the category of "quasi existence" (Ibid 27).

III – "Quasi existence" and the dualism of scheme and content

The quasi existing things of Dennett's intentional stance have no ontological significance in the traditional sense, even while they maintain their familiar conceptual position in our language. The term "mild realism" thus seems to inflate the status Dennett has in mind for beliefs and selves. It might be clearer to drop talk of realism about intentional concepts altogether. Dennett confines "realism," however, to a "natural ontological attitude" that has already excluded "opinion about the ultimate metaphysical status of physical things or abstract things (e.g., - electrons or centers of gravity)" (Ibid 30). The "natural ontological attitude" seems to be in keeping with Donald Davidson's rejection of the dualism of (linguistic; theoretical) scheme and (empirical) content. That dualism entails the doctrine that we can make sense of the idea of reality apart from our schemes, whether construed as unprocessed empirical data, uninterpreted experience, or the thing in itself. In this and following sections, I will suggest that, notwithstanding his exclusion of the ultimate status of things, Dennett's "mild realist" policy lists towards this untenable doctrine (Ibid 50).

The idea of existence coming in degrees or grades seems to depend on a non-relative idea of reality, however we characterise it. But a non-relative idea of reality does not readily support the idea of degrees of reality. In this respect, Dennett's decision to put aside the idea of ultimate reality begins to resemble a sceptic's or a relativist's effort to reject the idea of reality *per se*. The idea of reality provides traditional scepticism and conceptual relativism with a hidden epistemic leverage with which to bring into question the reality of the world that we seem to encounter. Traditional scepticism promotes blanket uncertainty by raising dire questions about whether our perceptual apparatus lets us experience the real world; various forms of explicit and unannounced relativism raise or suggest similar questions about whether our linguistic-conceptual schemes are systematically distorting reality, trapping us in our own parochial versions of reality, or missing reality altogether. In his early essay "On the Very Idea of a Conceptual Scheme" (1974), Davidson sought to show that conceptual relativism, in all its forms, loses its grip on us once we give up the dualism of scheme and content. Early in the essay he raises the paradox that we can't make sense of differing relative conceptual schemes unless "there is a common co-ordinate system on which to plot them." The paradox Davidson was intent on underscoring is that "the existence of a common system belies the [relativist's] claim of dramatic incomparability" (Davidson 184, 1974). Davidson is well known for rejecting both ideas, for arguing, on the basis of his views of language, against the idea of a common linguistic-conceptual system which just happens to capture ultimate reality, and against the idea that there are differing relative systems picturing different worlds. I discuss Davidson's view of language below;[1] for now we might consider that Dennett is not *so much*[2] committed to relativism construed as differing systems that each correctly picture the world, or differing worlds, but to a limited relativism in which not all relative schemes are equal, in which some schemes are more real than others. This limited relativism seems to be a presupposition of his view that various epistemic stances describe reality in *more* or *less* fundamental terms.

However, if Dennett is not tacitly banking on, while claiming to suppress, an idea of ultimate reality in contrast to relative reality, it seems odd that he should keep his view that some parts of reality are semi-real. It is not entirely clear how, without ultimate versus less-than-ultimate versions of reality, he can make sense of the idea of degrees of reality, as opposed to the traditional view that objects either exist or they do not exist. This point can be put in positive terms. If limited relativism, which tacitly requires some spectral version of ultimate reality, goes by the wayside, we are left with the ordinary reality which we encounter *with*, as opposed to mediated by, our language, and so we are left with the usual things and streams of things – from physical items to intentional things, beliefs, other selves, and so forth – which shape, and in the broadest and most innocuous sense are shaped by, that language.[3] On this view, no class of things featured within the totality of our language provides the smallest point on which to leverage some piece of reality into an ontologically higher status, or to relegate other pieces to a lesser quasi-status. We thus cannot maintain the idea of reality in degrees. Instead, if this is right, we are left with an indivisible ontology, with a reality in which we are fully immersed, and, not incidentally, with beliefs about reality which are either true or not true.

IV – Cartesian dualism and a false dilemma

Dennett suggests a motive for introducing different degrees or levels of existence, when he poses the following dilemma at the end of "Artifactual selves: a response to Lynne Rudder Baker":

> The only alternative on offer to this vision of an emergent, virtual [or quasi-] self
> is an extremely dubious dualism of real (*really* real, I guess you could say) selves
> in real bodies, an idea that is about as antithetical to a scientific understanding
> of mind as *élan vital* was to a scientific understanding of life.

So, for Dennett, the "mild realist" view of the self as a "virtual" or quasi existence is required if we wish to maintain "a scientific understanding of mind," which is under threat from Cartesian dualism, or some similar doctrine. This passage is surprising, because Dennett has, on other occasions, discussed a variety of non-dualist alternatives. In "Real Patterns" (1991), for example, he mentions five non-dualist views of the self: "industrial strength realism" (Jerry Fodor); "realism" (Davidson); "mild realism" (Dennett); "milder than mild irrealism" (Rorty), and eliminative materialism (Paul Churchland), which in effect denies the reality of selves (Dennett 30, 1991).

The most straightforward way to maintain "a scientific understanding of mind," or self, might be to deny the reality of these concepts. Dennett's disavowal of eliminative materialism, however, leaves him, he thinks (perhaps in an incautious moment), with a quandary in which either the self is a semi-real idealization *or* a window opens to the most extravagant non-scientific views of the self, e.g. "[t]he idea of a mind-thingy that goes to Heaven when somebody dies" (Dennett 333, 2013). As his survey of types of realism suggests, however, this is not a genuine quandary. It is not a quandary if realism has already been rescued from the false opposition of scheme-entranced relativists who deny the idea of objective reality and content-obsessed realists who posit some version of a Kantian thing in itself.

V – Pulling at threads

Dennett has missed an important aspect of Davidson's realism by misconstruing his fellow Quinian's[4] view of the role behaviour plays in making sense of language and the world. While enlisting Davidson against a theory which "takes beliefs to be things in the head," Dennett remarks that

> Davidson and I both like [Paul] Churchland's alternative idea of propositional-
> attitude statements as indirect "measurements" of a reality diffused in the
> behavioural dispositions of the brain (and body). We think beliefs are quite real
> enough to call real just so long as belief talk measures these complex behaviour-
> disposing organs as predictively as it does (Dennett 44-5, 1991).

By "belief talk" Dennett seems to mean talk which attributes belief to someone's behaviour, especially, I assume (to keep Dennett roughly on the same page as Davidson), the person's

linguistic behaviour. Davidson did take the view that understanding alien or idiosyncratic utterances (linguistic behaviour) involves, among other tasks, matching the utterances with the behaviour of the speaker. It is misleading, however, to say that an interpretation of these utterances predictively measures the behaviour of organs. This way of construing interpretation seems to be listing towards a reduction of these utterances to "the brain (and body)." Davidson's approach, however, avoids this construal of linguistic understanding, by keeping in sight three *inseparable* aspects: the utterances of a speaker's language; an interpreter making sense of those utterances and of the language to which they belong (perhaps to a passing language, a synthesis of the languages of interpreter and interpreted); and the familiar and unfamiliar objects of the world to which the speaker is more or less fluently referring. When Dennett, on behalf of Churchland, says that an interpretation (belief talk) of the utterances (beliefs) of a speaker are "indirect 'measurements' of a reality diffused in the behavioural dispositions of the brain (and body)," he seems to be operating in the vicinity of Davidson's triangulation of utterance-interpretation-world. Nevertheless, referring to interpretations as "indirect measurements," and as "diffused in the . . . dispositions of the brain (and body)" of the person to whose utterances the interpreter is ascribing belief, begins to dissolve one or more coordinates of the triangulation, for example interpretation and linguistic meaning, which are required by Davidson's account to make sense of utterances and beliefs. I won't press this point, as I'm not quite clear whether Dennett means that this diffusion of belief-talk amounts to an elimination of belief. But it is worth noting that Davidson has said that the concept of belief, alongside many other intentional concepts, is "essential to thought, and cannot be reduced to anything simpler or more fundamental" (Davidson 73, 2000), e.g. "the brain (and body)." It is also worth recalling that Davidson said, when distinguishing his views from Quine's, that he "despair[s] of behaviourism" (Davidson 231, 1979).

VI – Indeterminacy and evidence

After noting their (alleged) agreement, Dennett mentions that Davidson is more of a realist than he is about intentional concepts, such as belief and the self, and that he has "tracked down the source of this disagreement to a difference of opinion [they] have about the status of Quine's principle of indeterminacy of translation" (Dennett 46, 1991) and by extension indeterminacy of interpretation. The principle of indeterminacy of translation allows the possibility that two equally resourceful translation manuals of some foreign language, e.g. Jungle, might prescribe contrary and equally adequate interpretations of the same utterances in that language, Jungle. Indeterminacy of interpretation extends this possibility to the home language (Quine 48, 1992) so that two equally (but obviously not perfectly) justified interpretations of an utterance in the same language may be true.

In "Real Patterns," Dennett uses the following general example of indeterminacy, to explain why he is "less of a realist than Davidson" about beliefs:

> I see that there could be two different systems of belief attribution to an individual which differed *substantially* in what they attributed – even in yielding

substantially different predictions of the individual's future behaviour – and yet where no deeper fact of the matter could establish that one was a description of the individual's *real* beliefs and the other not (Dennett 49, 1991).

This quotation suggests two of the main ingredients of the scheme-content distinction: "different systems of belief attribution," and absence of a "deeper fact of the matter" that shows which system describes "the individual's *real* beliefs." All that is missing is an assertion that the "different systems" are untranslatable, which Dennett veers towards a few sentences later when he says that "[t]he rival theories [or systems] would not even agree on which parts of the world were [real] patterns and which were noise" (Ibid), and at the end of his paper when he claims that "Davidson has overlooked the possibility of two or more *conflicting* patterns being superimposed on the same data – a more radical indeterminacy . . . than he had supposed possible" (Ibid 51). The attribution of different patterns to the same data in itself is hardly surprising. A more dramatic possibility would be our inability to discern from the data and the rival systems of interpretation which data were encouraging acceptance of which system, the problem of inscrutability (or indeterminacy) of reference,[5] which Davidson spent decades thinking about in conversations with Quine and others.[6]

The drama of these problems begins to fade when we consider a point which Davidson has offered many times in essays directed at the related, dependent problem of conceptual relativism, namely that it is a mistake to conceive languages or systems of interpretation as fixed or impermeable over time. Conceiving of language as a finished structure, it is natural to suppose that the terms and references of one language might be untranslatable in another. Where translation fails, we can coherently raise cases of two similar systems of interpretation standing in permanent conflict, and in which the data on which they superimpose patterns might never let us see why we should prefer one system, or, within indeterminate or noisy areas of potential data, even what system we were dealing with. That would be quite a dramatic result, and if the data in question were the kind which normally encourages us to attribute particular beliefs to others, we might despair that we have a basis for belief attribution, and so a basis for our belief that others have beliefs. The drama would continue into the question of the self, as we would have no reason to suppose that we or they are selves if, in principle, *at some level* (invoking the idea of reality in degrees), the evidence we have for the existence of beliefs were to disappear. The idea of a self, after all, requires an array of concepts, and indispensable among these is the concept of belief.

As we have seen, for Dennett, beliefs, selves, and other intentional objects are theoretical fictions which we keep for purely pragmatic reasons. He thinks that we "always *can* have in principle a non-Intentional, mechanistic account of the entity" (Dennett 194, 1976) if we resort to a more basic epistemic strategy, based on physical evidence. We maintain the fiction of intentional objects only because the task of predicting the entity's behaviour would otherwise become extremely cumbersome. This view of the role of a basic epistemic strategy is the product of a mistake which Dennett seems to recognise, at least in part, as a mistake (Dennett 50, 1991). The mistake is to assume that there is a direct correspondence between some isolatable set of sentences of our languages, systems, or epistemic stances, and the available empirical evidence. The other mistake, which this assumption and Dennett's view of distinct

systems depends on, is tacitly to conceive language essentially as a fixed, impermeable, shared structure. This view of language leaves Dennett imagining rival finished structures which each enable some avenues of thought and close off others, and hence wondering sceptically at the possibility of translation between conflicting schemes.

Evidently, Dennett disavows this conception of language.[7] Yet something like it must be true if we can make sense of his rival systems of conflicting belief attribution or pattern superimposition. Dennett is stuck with such a conception if he wishes to maintain the drama of indeterminacy, because if the rival systems are not fixed, if their boundaries are permeable, then translation between rival schemes remains possible; and there is no reason to suppose that allowances for the terms and references of one scheme cannot be made by a radical interpreter starting with the other, as extensively as we care to imagine. Neither scheme need be construed as an epistemic, let alone ontological, trap. If an emissary from one scheme is sufficiently curious about the elusive objects of a rival scheme, radical interpretation will be required of her. At various points, she will need to take seriously aspects of behaviour and speech which, in her language, count as noise, or as grotesque or mistaken belief. Does that imply that she leaves behind her scheme or language on such occasions? Yes, if we mean nothing more than that she leaves behind her language conceived as a permanently fixed, shared structure; but if we have given up this idea of a language, it is no longer easy to make sense of the question.

Dennett stipulates that these rival schemes divide up the evidence differently; they superimpose *conflicting* patterns on the data. Does that arrangement impose a limit on radical interpretation? Not a theoretical limit; nor a task different from the one already engaging the radical interpreter as she makes sense of the references of another language by adjusting and expanding her own set of references, guided by, among many other beliefs, her beliefs about logic, which constrain her adjustments and, if she is charitable, prompt her to keep pressing her adjustments in the face of incommensurate schemes. It would only make sense to speak of an absolute limit on radical interpretation if we could make sense of pure evidence, or as Davidson might say "if we could give a non-linguistic characterization of reference," but as he added, and as Dennett perhaps accepts, "of this there seems to be no chance" (Davidson 221, 1977).

VII – Monstrous belief or myopia

Physical descriptions have no ontological priority over intentional descriptions if there is no chance to isolate reference from the rest of language, if reference in other words plays no independent role within the totality of our linguistic beliefs and theoretical activities. Without this ontological priority, Dennett's sceptical characterisation of beliefs and selves as semi-real things whose reality wanes as we descend to deeper levels of description, or take up more fundamental epistemic stances, is misplaced. At this point, should we concede that a "scientific understanding of mind" is out of reach, and be prepared to make room for the monstrous concepts of an ontology abandoned by reason? There may not be a general answer to this question. However, I would say that the semi-eliminativist, positivist inclination to conceive

such central intentional concepts as belief, value, and the self as a theorist's fiction tends to create its own monsters; it certainly creates a myopic conception of reality. That concern aside, the worry of a profligate ontology, unchecked by empirical evidence, facilitating production of grotesque or untenable beliefs – e.g., "a mind-thingy that goes to Heaven" – suggests a few misconceptions: that concepts which are inconsistent with empirical evidence need be entertained; that all non-analytic concepts are derived from sensory data; and perhaps most telling of all, that the concept of truth has no constraining role to play beyond scientific inquiries.

Rather than embark on distinct inquiries at this point (or attribute any of these views to Dennett), I will invoke Davidson's view of the role of truth (or error) in radical interpretation, taken as the means by which we expand our language – the interpretive activities of our historical linguistic horizons – when facing alien or incommensurate beliefs. Without a concept of error, there is a sense that we might believe anything. It would be more coherent, however, to say that without a concept of error or truth the concept of belief would dissolve, or, at least, we would have no basis for belief. Like radical interpretation, truth is forced on us if we wish to make sense of any meaningful description, descriptions that refer to our current commitments and those that refer to alien views and attachments which currently defy understanding. Each distinctly new candidate for belief that we make sense of when faced with an alien scheme presupposes adjustments in our language. We would not be in a position to call a belief "alien" or "incommensurate," let alone wonder whether it is grotesque or untenable, if we were not far along in this process. But even if we are far along in our understanding, nothing in our adjustments requires us to expand the reign of grotesqueness and error within our expanding view of things. We find ourselves in the opposite situation. Until we can discover or create plausible grounds within the totality of our beliefs and interpretations of reality, we are unable to begin to describe properly an alien worldview, let alone dismiss or applaud its more radically dubious commitments.

Welcoming new beliefs which cause us to adjust our core beliefs is thus an enormously involved task. I think it is implausible, and possibly unintelligible, to suppose that the task could be so fluent that we would ever find ourselves in a situation in which our core beliefs suddenly are systematically displaced by new conceptions, certainly not if truth is applied consistently across *all* beliefs in the new totality, and we maintain the idea of a self that persists through this change. This may be why Davidson once remarked that radical interpretation is as liable to "sharpen our perception of error" as it is to cause us to accept new beliefs.

VIII – Radical interpretation and self-transcendence

Only persons can apply radical interpretation, and the capacity to do so is essential to being a person. Lacking an ability to incorporate into her existing beliefs radically new beliefs, including those which disrupt who she is, no human would ever have developed into a person. So, without keeping in place the idea of persons as having an ability to engage in radical interpretation, it is hard to keep the idea of a person in sight, even though a disposition to

apply radical interpretation in the face of alien belief encourages a continual threat to the cognitive basis of a person's identity.

Nietzsche suggests an important aspect of this (existential) problem in his essay "On the Uses and Disadvantages of History for Life," where he struggles with an apparent tension between the unhistorical and historical modes of consciousness. "Unhistorical consciousness" and "historical consciousness" were terms he deployed to examine the idea that an individual (or a people or culture) might live more or less in the moment, enfolded in a closed (experiential-cognitive) horizon, or, alternatively, more among undigested and alien beliefs which continually require or tempt the individual to adjust her beliefs and thereby, if the attempt goes well, to expand her horizons. After introducing this theoretical scheme, Nietzsche invites us to consider that "the unhistorical and the historical are necessary in equal measure for the health of an individual, of a people and of a culture" (Nietzsche 63, 1997).

It is not clear how seriously we are meant to take this division, since persons would cease to be persons if their horizons were closed, or not informed by an inherently unstable field of background beliefs or presuppositions. I think Hans-Georg Gadamer was right when, reacting against Nietzsche's formulation, he concluded that the very idea of closed or separate horizons is unintelligible (Gadamer 304-5, 1989). Nietzsche's historical-unhistorical metaphor nevertheless might encourage us to reflect on practical quandaries, psychological and existential for instance, arising from the ongoing human activity of radical interpretation. As we digest the significance of core beliefs or encounter new beliefs which affect the core, our psychological resilience might be tested, and the kind of life in which we are immersed will in effect have changed, possibly independently of behavioural clues discernible to others. Dealing with psychological and existential quandaries is of course a normal part of the human condition. It is an inescapable part if we see the human condition as entailing a trajectory into personhood, and regard a capacity for radical interpretation as indispensable an aspect of personhood as beliefs, wishes, values, and other intentional states or activities.

Existential quandaries arising from radical interpretation might be germane to our inquiry about the ontology of the self. For they sometimes play an important role in our performance as radical interpreters, and that performance heavily determines how we become a particular self. The activity of radical interpretation of course need not play out in existential drama, and perhaps rarely does, at least not in ways that we appreciate well at the time. Gadamer suggests a less dramatic picture of the connection between radical interpretation and self-becoming when he describes the interpretive act of putting "ourselves in someone else's shoes," in the quite involved sense of becoming "aware of the otherness, the indissoluble individuality of the other person":

> Transposing ourselves [in this sense] consists neither in the empathy of one individual for another, nor in subordinating another person to our own standards; rather, it always involves rising to a higher universality that overcomes not only our own particularity but also that of the other (Ibid 305).

Gaining a new perspective, on Gadamer's account, requires the radical interpreter to overcome within herself some particular belief, aspect of sensibility, or interest in the world. In that sense, she must change herself. But we might wonder whether a new perspective or awareness of idiosyncratic features of the other person "always involves rising to a higher universality," and whether an interpreter's awareness of a radically new belief need bring a lasting change in her.

The durability of the change naturally depends on many factors. Since radical interpretation requires an effort to understand the alien belief as true, or to perceive why the belief is mistaken, the act or activity of radical interpretation may carry an existential risk, or a risk to an individual's way of perceiving and existing in the world. Understanding the truth or falsity of alien beliefs, after all, requires adjustments among an individual's own beliefs and linguistic commitments. If her awareness of the alien belief is secured at the cost of a revision of her beliefs which changes how she regards various points of interest in the world, or ramifies through her beliefs in a way that causes other adjustments among her beliefs and propensities, her person, prior to achieving a new totality of beliefs and interests, will have become unstable. But then neither her new nor her old system is ever finished, at least while she is alive and experiencing the world. So, with respect to the totality of her beliefs and values, her person in any event should be conceived as in flux.

If we reject the idea of persons bound by closed horizons, the proposition that persons are not stable entities might be trivial, even though there is a sense in which an individual might or might not practice radical interpretation when faced with alien beliefs or more complicated ways of seeing older beliefs, or when considering the risks or labour of an inquiry that makes significant demands on her intellect or peace of mind. An individual's *willingness and ability* to undertake the demands of radical interpretation suggests a predisposition in favour of self-overcoming, or of taking responsibility for the state of certain beliefs that define her outlook and ways of encountering the world, and for assimilating alien beliefs threatening or promising to change that outlook and disposition. Nietzsche characterised this capacity as a sign of strength of a person's "innermost nature" (Nietzsche 62-3). Possibly taking his historical-unhistorical distinction too seriously, he nevertheless claimed that "this is a universal law: a living thing can be healthy, strong and fruitful only when bounded by a horizon; if it is incapable of drawing a horizon around itself . . . it will pine away slowly or hasten to its timely end" (Ibid).

We can make sense of this claim if we abandon the incoherent idea of a *person* bounded by a horizon, and instead refer to the inherent limits of her ability or her inclination properly to assimilate certain new beliefs, of the practical limits of an actual radical interpreter's capacity to revise her beliefs and linguistic commitments and thereby to expand her worldview and language. In this way, we can keep Nietzsche's admiration for the capacity or "strength" of the radical interpreter more or less safe from his concern that the activity of radical interpretation undermines her prospects for staying "healthy, strong and fruitful" when her "horizons" are open to the alien. To the extent that she is unable to make sense of alien beliefs and proclivities, they will not become part of her expanded view of things, let alone of her disposition and values. Insofar as she *can* make sense, they have found a temporary place among the totality of her beliefs, and possibly in her disposition to act in certain ways. What should we say if her cognitive/spiritual health, strength or fruitfulness diminishes? Perhaps that

her capacity for self-examination is unduly limited, or that it has become impaired. Or that she is not resilient psychologically. The issue in any case is not that the "horizons" of her beliefs or values are unstable; nor that she would be a healthier, stronger and more fruitful person were she locked into a scheme from which she views and acts in the world. These are not issues we should try to make sense of, unless we can make sense of the idea of a closed system of beliefs and values, and of a *person* existing in such a system.

IX – Self-overcoming, language and consciousness

Radical interpretation is essential to being a person because self-overcoming is. It would seem that radical interpretation can, in some sense, begin to work beneath the surface of self-conscious reflection, e.g. when an individual changes core beliefs that affect how she perceives and acts in the world without explicit deliberation. An individual can also change as a result of self-examination, perhaps in "a dialogue with oneself" (Dennett 193, 1976), and it may be that these two kinds of activities cannot easily be disentangled.

In *The Gay Science*, Nietzsche ignores or overlooks this last thought when he distinguishes emphatically between features of the self which operate beneath consciousness to motivate action and the same features when they are, in some sense, translated into consciousness. Drawing on this distinction, he offers the surprising doctrine that "consciousness does not really belong to man's individual existence but rather to his social or herd nature":

> Our thoughts themselves are continually governed by the character of consciousness . . . and translated back into the perspective of the herd. Fundamentally, all our actions are altogether incomparably personal, unique, and infinitely individual . . . But as soon as we translate them into consciousness *they no longer seem to be* (Nietzsche 299, 1974).

It is not difficult to see the basis of this view; it stems from a fairly blunt version of the scheme-content distinction:

> Owing to the nature of *animal consciousness*, the world of which we can become conscious is only a surface- and sign-world, a world that is made common and meaner (Ibid 299-300).

Nietzsche says that his intent is not to characterise the distinction of (linguistic) scheme, through which we become conscious, and (distorted) content as metaphysical in character; [8] nevertheless, his account of consciousness bears all the hallmarks of the distinction:

> whatever becomes conscious *becomes* by the same token shallow, thin, relatively stupid, general, sign, herd signal; all becoming conscious involves a great and thorough corruption, falsification, reduction to superficialities, and generalization (Ibid).

Language and consciousness here have become a kind of display case in which we see "a world that is made common and meaner," which has been systematically distorted by language; and notwithstanding the facility with which the problem has been shown to us, we are evidently bound by the semantic (herd) conventions which present this distorted world, and by the lamentable fact that we are conscious.

It would take a bit of a procrustean manoeuver to bring Nietzsche's and Dennett's concept of the self into alignment. But if we bracket Nietzsche's proposal that the proper locus of the self is the unconscious, and take on board his view that a person occupies a "shallow, thin" version of herself in conscious states that are shaped by the influence of linguistic schemes, we are left with an idea of persons approximating Dennett's virtual self. Of course to keep this similarity we need to eliminate the idea of an underlying more authentic version of the self, which Nietzsche himself does on other occasions.[9] In place of that mysterious idea, we can refer to unconscious neural activities; these activities are not the self but the cause of the behavioural patterns to which, according to Dennett, we attribute the theoretical narrative fiction of "the self," an interpretive imposition that helps us predict future behavioural patterns.

Can we maintain the idea of a self that is capable of transcending itself if we construe the self as a theoretical construct? In his 1976 paper "Conditions of Personhood," Dennett takes this capacity of self-transcendence to be a necessary condition of personhood. And, in that paper, he says that to "be in a position to induce change in oneself," one must be capable of "reflective self-evaluation," taken as a form of "genuine self-consciousness" (Dennett 193, 1976). He then, as we have seen, pulls the rug out from this possibility:

> There is no objectively satisfiable sufficient condition for an entity's *really* having beliefs, and as we uncover apparent irrationality under an Intentional interpretation of an entity, our grounds for ascribing any beliefs at all wanes, especially when we have (what we always *can* have in principle) a non-Intentional, mechanistic account of the entity (Ibid).

If there are no objective grounds even for attributing beliefs to any entity, then the idea of an entity which can consciously reflect on its beliefs, choose to reconsider its core values, and thereby "to induce change in [itself,]" does begin to look like a fiction. Dennett concludes "Conditions of Personhood" by saying that "[w]hen such problems arise we cannot even tell in our own cases if we are persons" (Ibid).

Whereas Dennett concludes in "Conditions of Personhood" that we cannot account for persons, his current semi-realist view of persons as a theorist's narrative fiction falters because it cannot adequately account for the concept of self-transcendence. The aim of the fiction is to permit a theorist to organise, for the purpose of prediction, the kind of behavioural items that we associate with persons. The theorist, we are told, posits the fiction "in order to unify and make sense of an otherwise baffling complex collection of actions, utterances, fidgets, complaints, promises, and so forth, that make up a person" (Dennett 336, 2013). We are told, further, that the self thus posited is capable of "accumulating memories and devising plans and

expectations" (Ibid). Conceptually, a good deal is being presupposed here, more or less what we ordinarily presuppose when referring to an actual person. I would say that this theorist's posit is indistinguishable from an actual person, unless all the concepts being called into service – "actions, utterances . . . complaints, promises, and so forth" – are pseudo concepts, or mean something very different from our familiar concepts. That seems to be the case. We seem to be dealing with the "shallow, thin . . . surface- and sign-world" concepts which Nietzsche invokes when describing his herd self, with the added twist that the fictional self is not conscious. Or is it? If the theorist's self that Dennett describes *is* conscious, then his distinction between a theorist's fiction and an actual self disappears.

Being conscious in any case is not enough. To be a person one must, as Dennett says in his 1976 paper, be capable of the kind a self-conscious re-evaluation of beliefs and values which can induce us to change ourselves, our defining outlook, without which we would be incapable of understanding radically distinct utterances and points of view. Without this capacity for radical interpretation, we would not be able to make sense of the "baffling complex collection of actions, utterances . . . and [other phenomena], that [actually] make up a person." This is the endpoint of Dennett's account of the self. For if we allow that the complex collection of intentional phenomena and locus of self-evaluation and change with which we each seem to be intimately familiar exists in our own case, we have in effect accepted that persons are real and not merely a theorist's fiction, *which after all need not be posited, in the sense just described, for the self to exist.*

X – A sceptic's dogma

Does the idea of a self that exists beyond a theorist's construction re-open the door to a dualism of content and scheme? Only if that idea entails the idea of an unperceived, un-depicted self, or the self as a mysterious experiencing substratum. Obviously the content of a person's consciousness vanishes without some (self-) perceived, depicted form,[10] since there would otherwise be nothing of which to be conscious. Accommodating, and perhaps extending, the behaviourism which has beguiled Dennett, we might characterise the subjective contents of consciousness – beliefs, feelings, private wishes, etc. – as forms of inwardly perceived behaviour. This characterisation undermines a substantive subject-object distinction, since any inner phenomena of consciousness which a subject perceives has an accessible semantic structure which can be communicated to other persons, especially since these other persons routinely discover kindred phenomena among the contents of their own consciousness, and do so with language. This characterisation has another implication. If we give up the idea of "unstructured content" (Goodman 6) and can therefore no longer make sense of the idea of categorically private content, when we refer to phenomena of consciousness we may not need to distinguish between auto- and hetero-phenomenology, unless we wish to emphasise, for some reason, the initial proximity of content when describing or representing our own experience.

With the self thus exposed to public scrutiny and interpretation, we could naturally differ in our opinions about what inner and outer phenomena or behaviour, if any, count as a particular self. Even "in our own cases"? Individuals presumably are sometimes mistaken or uncertain about what configuration of their beliefs, values, and actions count as their essential self at any moment, or over a life, especially when their opinion comes into proximity or collides with the contrary, perhaps acutely perceptive, views of others. It is then not surprising that we should be unsure in particular instances, moment by moment or in later reflection, whether the identity conditions of an entity as variable as the self have been satisfied. But perhaps we should step back from an examination of (inner and outer) behaviour, from a direct appeal to the observable phenomena of the self. When expressing his doubt about whether we can "even tell in our own cases if we are persons," Dennett focussed on the elusive condition of self-transcendence. He was right to do so, as that condition of personhood presupposes other necessary conditions of personhood, such as rationality and language use. And he was right to doubt that we could know, using exclusively behaviourist methods, whether the condition of self-transcendence is ever satisfied. As he suggests, a sceptic might always pose contrary interpretations of what we observe, or question whether any interpretation of what counts as the self can gain a foothold. Scepticism along these lines, however, is problematic, since it involves isolating sentences purportedly counting as direct evidence and pressing the question of their correspondence, which is the problem of indeterminacy of reference which we considered earlier. I think we should take a broader approach to the issue of whether we ever satisfy the conditions of personhood, especially the elusive conditions of self-transcendence. The conditions of self-transcendence would seem to overlap with those of radical interpretation, and it may be that to be satisfied they require that radical interpretation can succeed. So, let us consider the prospects for these conditions together.

In particular instances, it is hard to know if an individual has demonstrated her ability radically to transcend her core beliefs and values, and therefore whether she has surmounted fundamental aspects of her identity. An individual might, at some point of self-reflection, think that she has satisfied a condition that reveals her personhood, by bringing into question and changing some aspect of her beliefs. But every direct assurance she can offer can be brought into question by a sceptic, who can reasonably persist in asking what beliefs or values are core to her identity, which interpretive scheme correctly identifies her (Dennett's worry or challenge). We might also wonder whether her potentially distorting language allows her to think beyond its confines, to let her assess radically her values in the face of values from which they fundamentally differ, or merely leaves her stymied among arbitrary configurations of her old beliefs. Confining herself to the (inscrutable, indeterminate) evidence at hand, it may seem that she will never be in a position to prove that she is not locked into her language and beliefs, and so it may seem that she can never gain a foothold against these kinds of sceptical queries.

The first of these challenges can be addressed more straightforwardly than the second. The claim that she is essentially stuck in her system of beliefs comes into conflict with the evidence of her memory, and her comparative judgement of her previous and present beliefs and values. That challenge should leave her unimpressed if she has no reason to distrust her memory and her judgement that she has significantly changed her beliefs. A more general challenge, that her

language is an enclosed system beyond which lies meaning inaccessible to her, need not undermine her assurance that *sometimes* she is capable of surmounting her worldview, her core beliefs and values and ways of seeing reality; it would merely imply, *if successful*, that she might develop in some ways and not in others. But as we saw in section 5, it is not clear how such a challenge could be successful in a general sense; it would require us to picture language as an impermeable finished structure, impervious to radical interpretation, to actual interpreters, no matter how resourceful they might be.

Dennett's particular worry, or challenge, that *someone else* might attribute to her beliefs and other conditions of personhood from an incommensurable rival scheme is more interesting, though it is also burdened by the idea of an impermeable language. Moreover, as we saw in section 5, buying into this challenge requires us not only to impose a theoretical limit on radical interpretation *per se*, but to accept the idea of isolated content or pure evidence, and therefore requires us, as Davidson would say, to "give a non-linguistic characterization of reference." But Dennett's worry cannot easily be thrown over, as it presents the plausible view that others (lovers, neuroscientists, etc.) might, at times, understand aspects of you better than you understand yourself, and so be in a better position than you are to say what you amount to, or who essentially you are. Even divested of the idea of incommensurability, this worry might weaken the prospects of personhood, at least in particular instances, perhaps to the extent that personhood requires that you are able to satisfy the conditions of self-transcendence; for I suppose, in principle, others could sometimes be in a better position to change you than you are yourself – and an entity that is *exclusively* the outcome of the manipulations of others would lack the degree of autonomy required to count as a person.

XI – Indeterminacy and openness

If self-transcendence is an essential aspect of personhood, as it seems to be in our principal or most influential traditions,[11] the idea of a determinate, substance-like self, which features in some of these same traditions, would not make sense. If we regard self-transcendence as an essential aspect of personhood, then persons, if they exist, are in flux, fixed it would seem only in relation to the development of their beliefs, values, and deeds. Is there anything *essential* about a self thus conceived? It would seem not with respect to any collection of beliefs, values, or deeds, considered in isolation and fixed in time. If self-transcendence is a condition of personhood, then a person's current beliefs and values would seem to be essential only through a wider set of relationships, which minimally, I would say, include the beliefs and values from which these current beliefs have grown, and whatever deeds this evolving set of beliefs and values has motivated or prevented.

This stream of events is consistent with a succession of selves, and there seems to be no method for deciding divisions among the prospective selves that can be attributed to our bodies over the most varied or uneventful lives, e.g., for the sake of ascribing responsibility on an ethical basis for various deeds and misdeeds. As dramatic as this outcome may be, it does not entail an annihilating indeterminacy that extinguishes the very possibility of identity

conditions for the self. Contrary attributions of features which purport to define a person suggest a continuously uncertain area of inquiry, and of self-inquiry, in which competing narrative accounts can readily be brought into collision. Does this situation undermine the idea of a coherent self? Contrary inquiries and self-inquiries might misconstrue a particular self, and in some measure likely all do. Yet the tensions which sponsor that likelihood also underwrite the very idea of a person. Collisions of perspective about a self, whether originating to some extent within the self in question, or prompted by the queries, interpretations, influence or actions of others, play a *constitutive* role in the development of a person, by influencing the continuous re-evaluation of beliefs, values, deeds, reflection on deeds, and so forth, that underwrite self-inquiry and self-transcendence, and therefore the (ongoing) possibility of a person.

Of course collisions, or exchanges, of perspectives between a self and another, or others, may contribute to forms of self-transcendence grounded in reality, *or*, alternatively, to arbitrary, self-deceptive or delusional forms of self-transcendence, though the latter possibility might impose a limit on how the word "transcendence" applies, and encourage us to think, in some cases, of personal change in a neutral sense, or to think of a person degenerating in some way. Perhaps only an individual who reflects on such collisions within herself *as a radical interpreter*, motivated by charity and therefore constrained by the ideal of truth, or by a willingness to perceive error among her beliefs and values, can properly embody instances of *self-transcendence*. This may tie the concept too closely to rational agency, another condition of personhood. While persons must be construed as agents, they are not always busy satisfying the conditions of agency, deliberating and making decisions on a rational basis. Exactly how persons change, for good or ill, is something of a mystery. As I mentioned at the outset of section 8, self-transcendence (self-overcoming) does not seem to be a function of agency alone; sometimes it is founded in rational deliberation and sometimes not, or only distantly or obliquely.

XII – Self-transcendence and *physis*

A fairly telling observation to direct against the idea of virtual persons is that persons have bodies. A virtual person is a theoretical construct, whereas we ordinarily regard a person as physically embodied. It would be quite hard to disentangle how we encounter a particular person, let alone our interactions with her, from her body, e.g. her facial expressions or the movement of her limbs. The beliefs, values, desires, actions, and so forth, that we attribute to a particular physical entity, a potential person, might be too chaotic, or our intuitive powers might be too impoverished, to lead us to posit a person. But these intentional objects are as bound to physical reality, are as much part of nature, however we construe nature, as the physical entity to which we attribute them; and if we suspect that the chaos is minimal enough to meet the approximate criteria of personhood, the person whom we might then be justified in positing is as entirely physical as the observable physical parts of her animating body. If we accept this assumption, it does not follow that we should reduce beliefs, values, desires and other intentional objects, right up to persons, to assemblages of neurons or to areas in the

brain. Intentional objects, certainly beliefs, also have a content or meaning, and so, short of eliminating them, they are not reducible to neurons or brain parts. Instead we might keep our inquiry into the nature of belief, the self, and so on, open, maintaining a dualism of explanation, as Davidson suggests, while regarding beliefs, selves and physical objects as part of the same ultimate nature, whatever that might be.[12]

Regardless of how we settle such ambitious inquiries, we should see beliefs and, obviously, values, wishes, desires, and persons, as inextricably bound to *passions*, which are inextricably tied to the body. This would make less plausible the view that self-transcendence is simply, let alone exclusively, a rational enterprise, which suggests an absurdly narrow view of persons. Are we merely appealing to Hume's famous quasi-irrationalist, quasi-emotivist dictum that reason is and ought to be regarded as the slave of the passions? I think the distinction is more elusive. The following question about our relationship to truth and belief suggests how we can regard rational belief as inseparable from the passions without seeing reason as subordinate: Does reason become the slave of the passions when we enlist reasons to confirm a belief, or is it that our passions are governed by reason when we come to accept a belief as true on rational grounds? This is not a good question if there is no cleavage *within us* between the operation of reason and passion, if beliefs and related dispositions and operations have a cognitive *and* physiological dimension. We are constitutionally *impelled* to prefer true beliefs over false, and it is *rational* to do so, at least if truth is essentially tied to reason.[13] To be rational in this sense is to be grounded in the passions, which reflect both our cognitive and physical nature.

There is of course a straightforward sense in which reason and the passions can come into opposition among our beliefs, though this sense refers to a conflict of desires in which some desires fall more on the side of rational norms and knowledge than others; if beliefs are inextricably constituted by reason and passion, the opposition between these would not be categorical. We could still reasonably say that there are times when the opposition is resolved in favour of reason. We might, for instance, offer instances of self-transcendence in which rational passions prevail over comparatively less rational or irrational passions. Naturally we could offer instances of the reverse trajectory. Indeed it may seem mysterious why more rather than less rational beliefs should ever prevail, at least beyond the calculative dictates of survival and utility. While it is hard (perhaps incoherent) to imagine anyone choosing and not merely professing beliefs which they are convinced are untrue, the activity of examining our core existential beliefs and seeking ways to falsify their hold on us does not seem to come naturally. Beliefs descend on us without our consent, and those which we acquire from a standpoint of self-reflective thought and inquiry are enabled by the rest of our beliefs, which are mostly the unexamined product of a tradition, or of indoctrination. Further, the remarkable transformation of a human into a person who can examine her beliefs very far at all is the outcome of lengthy historical developments of a variety of civic institutions and practices, and the development of elaborate forms of knowledge and methods of inquiry, all based on the platform of language. The idea of persons as self-conscious agents capable of autonomous change thus might seem to refer to a highly artificial construction after all, a non-natural product of culture and language.

This is Nietzsche's position in *The Gay Science*, where, as we have seen, he argues that the social and linguistic nature of self-consciousness means that "the world of which we can become conscious is only a surface- and sign-world" in which the conscious self "*becomes* by the same token shallow, thin, relatively stupid, general, sign, herd signal." His view of the self elsewhere tends to undermine this characterisation, for example when he presents, in the Second Essay of *On the Genealogy of Morals*, his famous hypothesis of the origin of the conscience as based in the "internalization" (repression; sublimation) of desire:

> All instincts that do not discharge themselves outwardly *turn inward* – this is what I call the internalization of man: thus it was that man first developed what was later called his "soul." This entire inner world, originally as thin as if it were stretched between two membranes, expanded and extended itself, acquired depth, breadth, and height, in the same measures as outward discharge was *inhibited* (Nietzsche sec 16, 1967).

Here Nietzsche takes the view that a fairly significant change in "man's" *nature* occurred, at some point in his transition into civil society, which prepared the way for human intellectual capacities to develop as elaborately as they have. He no longer seems to regard the "entire inner world" (consciousness) of humans who undergo this change as merely "a surface- and sign-world" in which the conscious self "becomes . . . shallow, thin, relatively stupid," and so on. Instead he suggests that the elaborate intellectual capacities of persons are both a product of history and *thoroughly* continuous with the nature of the self.

XIII – Persons, thought, and the world

How we construe this last point, which in some sense should seem unremarkable, may determine whether or not we should accept the view that the spiritual-mental activities of persons are theoretical add-ons, fictions spun out by our brains to account for the behaviour of entities which are not *really* persons, or as emergent states, the causally inert endpoint of a one-directional causal order. Both views, I think, are mistaken, but the theorist's fiction view is mistaken in two respects. It shares with the emergent-state theory of consciousness and persons the defect of not being able to account for how persons can act in, or on, the world. More problematically, the theorist's fiction theory replaces a metaphysical theory in which subjects and objects are ontologically distinct with a theory that only postulates objects. The presupposition of this theory is that there is no subjective aspect of the world, *really*. As we have seen, Dennett recommends that we postulate persons (subjects) and their various intentional states only in order to facilitate less cumbersome predictions of the physical things to which we attribute these conceits. For practical reasons, on Dennett's view, it would be premature to eliminate the subjective dimension of reality, but not in principle impossible. This is one way to abandon subject-object dualism, a familiar enough option these days. But it overlooks the most compelling reason for abandoning the doctrine, which, as we have seen, is this: that there is no way to disentangle our theories and language from the reality we seek to understand.

This motive for rejecting the subject-objection distinction undermines naïve physical realism, but the reason for rejecting the distinction lends no support to a subjectivist theory of reality. Rejection of the distinction involves abandoning a metaphysical picture in which reality contains two kinds of things, but it does not entail preserving half the distinction, either by adopting naïve realism or subjectivism. The holistic view of language and reference that motivates rejection of the distinction undermines replacement doctrines that would categorically dissolve the tension that we subjects commonly experience when we encounter a world of objects, and when we encounter ourselves and other subjects within, as part of, this same world. Quine famously offered, in "Two Dogmas of Empiricism," a critique of the doctrine that we can confirm descriptions of objects in the world by a direct, unmediated appeal to evidence, a conclusion which arguably Davidson took more seriously than his former teacher. This kind of appeal is doomed if we are unable to isolate our references to these objects, and our references to the kinds of evidence being invoked on their behalf, from the rest of our language and theories. That we could do so, while retaining the meaning of these references, is hard to conceive. A similar point can be made with respect to references to things that appear to be distinctly subjective: beliefs, desires, wishes, etc. These, as much as quarks or quantum fields, cannot be made sense of in isolation from the entirety of our language and theories, even if particular inquiries within our understanding of reality have adopted terminologies and models in which few of our familiar (common sense and traditional) terms show up. We can therefore reject the kind of direct appeals to subjective phenomena that would uphold a purely subjectivist theory of reality for the same reason that we reject a naïve empiricist's direct appeal to evidence.

This basis for rejecting the dualism of subject and object should encourage us to maintain the pre-scientific view that the world is much like we experience it, full of objects and experiencing subjects. As subjects, we commonly experience a tension between our intentional subjective experience and the objects we perceive or infer, even if an intelligible understanding of the world presupposes an integration of subject and object. This tension can be quite acute when we refer to *persons*, especially if we emphasise experiences in which a person feels "enslaved" by, or alienated from, her body, or vice versa, as the Body in Marvell's poem "A Dialogue Between the Soul and Body" asks us to imagine:

> O who shall me deliver whole,
> From bonds of this tyrannic Soul? (Marvell 1978)

The impossibility of the Body's query/complaint – beyond the absurdity *that* a body queries or complains – suggests the futility of trying to collapse the tension we experience in ourselves, and in relation to other subjects, into one side or other of the dualism the poem brings into question. The Soul's complementary complaint that it is "hung up . . . in chains/ Of nerves, and arteries, and veins," reinforces the suggestion from the other side, making it clear that any sympathies we might attach to the Body in its plight, or the Soul in its, will be left hanging on positions dissolved by the parody of the dialogue.

Nietzsche's theory of repression pictures the conscience as a "tyrannic Soul," but it construes the "soul" as a collection and *transfiguration* of human passions, as thoroughly part

of a unified and "improved *physis*," not as a distinct kind of thing which eliminates passions and thereby transcends the body. At the end of "On the Uses and Disadvantages of History for Life," Nietzsche extends this thought to the entire world of intentional objects, when he proposes that we view Greek culture:

> as a new and improved *physis*, without inner and outer, without dissimulation and convention, culture as a unanimity of life, thought, appearance and will (Nietzsche 1997).

There is sense in which a culture, like a worldview, does not exist; conceived as a distinct object, it is an abstraction whose intended reference dissolves under scrutiny. There seem to be particular kinds of intentional experience only available to individuals who have developed within the *cultural forms of life* of a *society* or a *tradition*. But these are kindred abstractions. Along with the sort of things – practices, institutions, conventions – which populate cultures, societies and traditions, these abstractions have no independent meaning. A virtue of Nietzsche's proposal, it seems to me, is that it encourages us to think primarily of the *individuals* from whose intentional experiences we extrapolate, in order to think *concretely* about culture, society, tradition, convention, and so on. Obviously individuals are influenced by their cultural and social circumstances, but if we reject the idea of fixed cultures, with "inner and outer" boundaries, we could then more simply say that individuals, in their development, communicate in various interesting ways with one another, sharing the content of their intentions and beliefs, and, in so doing, giving meaning to social practices and institutions. Moreover, if, following Nietzsche, we see this significance sharing "inner world" of the individual as transfigured *physis*, we would not merely be positing further abstractions, theoretical fictions or emergent sets of purely mental properties. We would instead find ourselves near Davidson's view that concepts such as "intention, belief, desire, and so on . . . are essential to thought, and cannot be reduced to anything simpler or more fundamental" (Davidson 73, 2000).

This view is a corollary of Davidson's general approach to language and ontology, which cannot be divorced from the reality of subjects, actual language speakers. As we saw in section 5, Davidson offers a theory of language in which the ability to speak a natural language, and therefore understand reality, requires subjects in two connected respects: to say things about reality and to interpret what is said. Utterances and interpretations, in Davidson's theory of language, are co-ordinated (triangulated) with the world, and therefore with the truth of utterances and interpretations, or the possibility of their error. Could we simplify Davidson's theory, by subtracting the intersubjective requirement of language? Even if we require a triangulation of *utterances*, *interpretations* and *the world* to explain linguistic communication, do we need to keep subjects? For example, might not computers play the same role? In some sense, after all, computers perform both these functions: they generate statements and interpretations. This is an intriguing possibility, and becoming clearer about the conditions for making utterances and interpretations might provide a baseline for considering the degree or sense that we attribute thought to computers. The conditions for generating an utterance or an interpretation, however, make the prospects for their satisfaction formidable. For utterances and interpretations that communicate and are understood entail thought, and if Davidson is

right to say that concepts such as "intention, belief, desire, and so on . . . are essential to thought," the prospects for intelligibly attributing thought to computers are negligible. Accepting Davidson's contention that intentional objects "are essential to thought" then makes exceedingly implausible the view that linguistic communication could occur without actual subjects.

The world and the concept of truth are also "essential to thought." And since "*all* these concepts (and more)" (Ibid) are essentially interrelated in the activity of language, it is natural, if Davidson is right, to think of subjects as essentially part of the fabric of the world. That thought follows from the principal motive we have for abandoning the dualism of subject and object. It might be more palatable to say that subjects are an essential part of the world insofar as creatures capable of thought have evolved. I think that must be right.[14] It appears that such creatures have evolved. If some of these creatures have evolved an ability to understand utterances that require of them the kind of self-reflective thought and degree of autonomy assumed by radical interpretation, and on occasion show that ability, then, from as objective a standpoint as can be imagined, it is clear that persons sometimes appear in the world.

Works Cited

Bourke, Michael. "Truth and Fictional Meaning," *Ontology Studies*, Volume 12, 2012.

Dennett, Daniel. "Are We Free?" Neuroscience gives the wrong answer," *Prospect*, October 16, 2014 (November issue).
-----. "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* (2013).
-----. "The Role of Language in Intelligence," *What Is Intelligence* (Cambridge, 1994), ed. Jean Khalfa.
-----. "Real Patterns," *The Journal of Philosophy*, Volume 88, Issue 1 (Jan., 1991).
-----. "Conditions of Personhood," *The Identities of Persons* (1976), ed. Amélie Oksenberg Rorty.

Davidson, Donald. "Truth Rehabilitated," *Rorty and His Critics* (Blackwell, 2000), ed. Robert B. Brandom.
-----. "Seeing Through Language," Thought and Language (Cambridge, 1997), ed. John Preston.
-----. "Meaning, Truth and Evidence" (1990), *Perspectives on Quine*, eds. Robert Barratt and Roger Gibson.
-----. "The Inscrutability of Reference" (1979), *Inquiries into Truth and Interpretation* (Oxford, 1984).
-----. "Reality Without Reference" (1977), *Inquiries into Truth and Interpretation* (Oxford, 1984).

-----. "On the Very Idea of a Conceptual Scheme" (1974), *Inquiries into Truth and Interpretation* (Oxford, 1984).

Gadamer, Hans-Georg. *Truth and Method*, Second, Revised Edition (Crossroad, 1989), trans. Joel Weinsheimer and Donald G. Marshal.

Goodman, Nelson. *Ways of Worldmaking* (Hackett, 1978).

Hume, David. *A Treatise of Human Nature*, Second Edition (Oxford, 1978), ed. L. A. Selby-Bigge.

Marvell, Andrew. "A Dialogue Between the Soul and the Body," anthologised in *George Herbert and the Seventeenth Century Religious Poets* (Norton, 1978), ed. Mario A. Di Cesare, 100-101.

Morris, Michael. *The Good and the True* (New York: Oxford University Press, 1992).

Nietzsche, Friedrich. "On the Uses and Disadvantages of History for Life," *Untimely Meditations* (Cambridge, 1997), trans. R. J. Hollingdale.
-----. "Schopenhauer as Educator," *Untimely Meditations* (Cambridge, 1997), trans. R. J. Hollingdale.
-----. *The Gay Science* (Vintage, 1974), trans. Walter Kaufmann. (Cambridge, 1997), trans. R. J. Hollingdale.
-----. *On the Genealogy of Morals* (Vintage, 1969), trans. Walter Kaufmann.

Quine, W. V. *Pursuit of Truth* (Harvard, 1992).
-----. "Two Dogmas of Empiricism," *From a Logical Point of View* (Harvard, 1980).

---

[1] I discuss Davidson' view of language at greater length, mainly as it relates to the troubling issue of fictional meaning, in "Truth and Fictional Meaning," *Ontology Studies*, Volume 12 (2012), pp. 389-406.

[2] As we will see in section 5, Dennett is *to some extent* committed to this idea.

[3] In "Seeing Through Language," (1997), Davidson uses the expression seeing *with language*, as opposed to *through or mediated by language*, to maintain his earlier rejection of a separation of (linguistic) scheme and (empirical) content.

[4] Davidson himself rejects this characterisation. See Ernest Lepore's "An Interview with Donald Davidson" (online interview), p. 15.

[5] Quine states a preference for "indeterminacy of reference" in chapter 20 of *Pursuit of Truth* (1992), 50.

[6] E.g., in the "The Inscrutability of Reference" (1979), *Inquiries into Truth and Interpretation*, and in "Meaning, Truth and Evidence" (1990), *Perspectives on Quine*.

[7] Daniel Dennett, "The Role of Language in Intelligence," *What is Intelligence?* (1994). In this paper, Dennett seems to view language more as a highly versatile cognitive tool than a pre-established system, and argues against the thesis that language is a *closed* system or likely entails "'cognitive closure' closure to some topics of inquiry."

[8] He characterises "the opposition of subject and object" as an epistemological distinction at the end of section 354 of *The Gay Science*.

[9] E.g., in section 1 of "Schopenhauer as Educator," *Untimely Meditations*.

[10] I have borrowed Nelson Goodman's phenomenological expression "content vanishes without form," from *Ways of Worldmaking* (Hackett, 1978), 6.

[11] I mean within Enlightenment institutes and the antecedent religious and philosophical traditions from which they borrowed the implied concepts of agency and personhood.

[12] Alternatively, we might question whether the best, most subtle, way to reject a dualism of matter and mind is to eliminate either. A point of Davidson's rejection of scheme and content that interpreters often miss, e.g. by thinking that schemes simply construct content, is that neither part of the distinction is forsaken. Similarly, perhaps we should refuse to forsake mind when we embrace metaphysical materialism, especially since it might not be coherent to think of a physical reality outside a logical structure.

[13] This point is inspired by Michael Morris' argument that truth is an intrinsic value in *The Good and the True* (New York: Oxford University Press, 1992), 9.2.

[14] The claim that subjects are essential does not imply that creatures capable of propositional thought needed to evolve, since the requirement of subjects to think about particular objects does not mean that objects could not exist without being thought about, which would be egregiously at odds with common sense. The requirement is merely a requirement for the creatures who postulate such objects. It constrains these creatures to postulate only those objects which can find a place in their theories, which is all that is needed to maintain the interrelationship of object and subject(s), world and theory. The requirement, in other words, amounts to the banal point that we can only include in our theories objects that are *susceptible* of being described, which does not preclude the possibility of objects existing in a world without subjects or existing in this world without being detected.