THE DISCONNECTION THESIS

By David Roden,
Department of Philosophy, The Open University

Forthcoming in Amnon H. Eden, James H. Moor, Johnny Søraker, Eric Steinhart (eds*.) The Singularity Hypothesis: A Scientific and Philosophical Assessment*, The Frontiers Collection, Heidelberg, Germany: Springer-Verlag 2013.

In this essay I claim that Vinge's idea of a technologically led intelligence explosion is philosophically important because it requires us to consider the prospect of a posthuman condition succeeding the human one. What is the "humanity" to which the posthuman is "post"? Does the possibility of a posthumanity presuppose that there is a 'human essence', or is there some other way of conceiving the human-posthuman difference? I argue that the difference should be conceived as an emergent disconnection between individuals, not in terms of the presence or lack of essential properties. I also suggest that these individuals should not be conceived in narrow biological terms but in "wide" terms permitting biological, cultural and technological relations of descent between human and posthuman. Finally, I consider the ethical implications of this metaphysics If, as I claim, the posthuman difference is not one between kinds but emerges diachronically between individuals, we cannot specify its nature a priori but only a posteriori. The only way to evaluate the posthuman condition would be to witness the emergence of posthumans. The implications of this are somewhat paradoxical. We are not currently in a position to evaluate the posthuman condition. Since posthumans could result from some iteration of our current technical activity, we have an interest in understanding what they might be like. It follows that we have an interest in making or becoming posthumans.

1. The Posthuman Impasse

In a 1993 article "The Coming Technological Singularity: How to survive in the posthuman era" the computer scientist Vernor Vinge argued that the invention of a technology for creating entities with greater than human intelligence would lead to the end of human dominion of the planet and the beginning of a posthuman era dominated by intelligences vastly greater than ours (Vinge 1993).

According to Vinge, this point could be reached via recursive improvements in the technology. If humans or human-equivalent intelligences could use the technology to create superhuman intelligences the resultant entities could make even more intelligent entities, and so on. Thus a technology for intelligence creation or intelligence amplification would constitute a singular point or "singularity" beyond which the level of mentation on this planet might increase exponentially and without limit.

The form of this technology is unimportant for Vinge's argument. It could be a powerful cognitive enhancement technique, a revolution in machine intelligence or synthetic life, or some as yet unenvisaged process. However, the technology needs to be "extendible" in as much that improving it yields corresponding increases in the intelligence produced. Our only current means of producing human-equivalent intelligence is non-extendible: "If we have better sex . . . it does not follow that our babies will be geniuses" (Chalmers 2010: 18).

The "posthuman" minds that would result from this "intelligence explosion" could be so vast, according to Vinge, that we have no models for their transformative potential. The

best we can do to grasp the significance of this "transcendental event", he claims, is to draw analogies with an earlier revolution in intelligence: the emergence of posthuman minds would be as much a step-change in the development of life on earth as the "The rise of humankind".

Vinge's singularity hypothesis – the claim that intelligence-making technology would generate posthuman intelligence by recursive improvement – is practically and philosophically important. If it is true and its preconditions feasible, its importance may outweigh other political and environmental concerns for these are predicated on human invariants such as biological embodiment, which may not obtain following a singularity.

However, even if a singularity is not technically possible – or not imminent – the Singularity Hypothesis (SH) *still* raises a troubling issue concerning our capacity to evaluate the long-run consequences of our technical activity in areas such as the NBIC technologies (Nanotechnology, Biotechnology, Information Technology, and Cognitive Science). This is because Vinge's prognosis presupposes a weaker, more general claim to the effect that our technical activity in NBIC areas or similar might generate forms of life which might be significantly alien or "other" to ours. I refer to this more general thesis as "Speculative Posthumanism".

If we assume Speculative Posthumanism it seems we can adopt either of two policies towards the posthuman prospect. Firstly, we can *account* for it: that is, assess the ethical implications of contributing to the creation of posthumans through our current technological activities.

However, Vinge's scenario gives us reasons for thinking that the differences between humans and posthumans could be so great as to render accounting impossible or problematic in the cases that matter. The differences stressed in Vinge's essay are cognitive: posthumans might be so much smarter than humans that we could not understand their thoughts or anticipate the transformative effects of posthuman technology. There might be other very radical differences. Posthumans might have experiences so different from ours that we cannot envisage what living a posthuman life would be like, let alone whether it would be worthwhile or worthless one.

For this reason, we may just opt to *discount* the possibility of posthumanity when considering the implications of our technological activity: considering only its implications for humans or for their modestly enhanced transhuman cousins. We can refer to the latter using Ray Kurzweil coinage "MOSH": Mostly Original Substrate Human (Agar 2010: 41-20).

However, humans and MOSH's have a prima facie duty to evaluate the outcomes of their technical activities of these differences with a view to maximizing the chances of achieving the good posthuman outcomes or, at least, avoiding the bad ones. It is, after all, their actions and their technologies that will antecede a posthuman difference-maker such as a singularity while the stakes for humans and MOSH's will be very great indeed.

From the human/MOSH point of view some posthuman dispensations might be transcendently good. Others could lead to a very rapid extinction of all humans and MOSH's, or something even worse. Charles Stross' novel *Accelerando* envisages human and MOSH social systems being superseded by Economics 2.0: a resource allocation system in which supply and demand relationships are computed too rapidly for those burdened by a "narrative chain" of personal consciousness to keep up. Under Economics

2.0 first person subjectivity is replaced "with a journal file of bid/request transactions" between autonomous software agents, while inhabited planets are pulverized and converted to more "productive" ends (Stross 2006: 177).

This post-singularity scenario is depicted as comically dreadful in Stross' novel. It is bad for humans and for their souped-up transhuman offspring who prove equally redundant amid such virulent efficiency. However, as the world-builder of *Accelerando*'s fictional posthuman future, Stross is able to stipulate the moral character of Economics 2.0. If we were confronted with posthumans, things might not be so easy. We cannot assume, for example, that a posthuman world lacking humans would be worse than one with humans but no posthumans. If posthumans were as unlike humans as humans are unlike non-human primates, a fair evaluation of their kinds of life might be beyond us.

Thus *accounting* for our contribution to making posthumans seems obligatory but may also be impossible with radically alien posthumans, while discounting our contribution is irresponsible. We can call this double bind: "the posthuman impasse".

If the impasse is real rather than apparent, then there may be no principles by which to assess the most significant and disruptive long-term outcomes of current developments in NBIC (and related) technologies.

One might try to circumvent the impasse by casting doubt on Speculative Posthumanism. It is conceivable that further developments in technology, on this planet at least, will never contribute to the emergence of significantly nonhuman forms of life.

However, Speculative Posthumanism is a weaker claim than SH and thus more plausible. Vinge's essay specifies one recipe for generating posthumans. But there might be posthuman difference-makers that do not require recursive self-improvement (we will consider some of these in due course). Moreover, we know that Darwinian natural selection has generated novel forms of life in the evolutionary past since humans are one such. Since there seems to be nothing special about the period of terrestrial history in which we live it seems hard to credit that comparable novelty resulting from some combination of biological or technological factors might not occur in the future.

Is there any way round the impasse that is compatible with Speculative Posthumanism? I will argue that there is, though some ethicists may prefer the *discounting* option to my proposal. However, to understand how the impasse can be avoided we must consider what Speculative Posthumanism entails in more detail.

As a first step towards this clarification, I will gloss the speculative posthumanist claim as the schematic possibility claim SP:

> (SP) *Descendants of current humans could cease to be human by virtue of a history of technical alteration*.

SP has notable features which, when fully explicated, can contribute to a coherent philosophical account of posthumanity.

Firstly, the SP schema defines posthumanity as the result of a process of technical *alteration*. Value-laden terms such as "enhancement" or "augmentation" which are more commonly used in debates about transhumanism and posthumanism are avoided. I shall

explain and justify this formulation in Section 2.

Secondly, it represents the relationship between humans and posthumans as a historical successor relation: *descent*. "Descent" is used in a "wide" sense insofar as qualifying entities might include our biological descendants or beings resulting from purely technical mediators (e.g., artificial intelligences, synthetic life-forms, or uploaded minds). The concept of Wide Descent will be further explained in Section 3.

Wide Descent also bears on one of the harder problems confronting a general account of the posthuman: what renders posthumans *nonhuman*? Is Speculative Posthumanism committed to a "human" or MOSH *essence* which all posthumans lack, or are there other ways of conceiving the difference?

I will argue that the account of Wide Descent, together with more general metaphysical considerations, militates against essentialism. I will propose, instead, that human-posthuman difference be understood as a concrete *disconnection* between individuals rather than as an abstract relation between essences or kinds. This anti-essentialist model will allow us to specify the circumstances under which accounting would be possible.

2. Value Neutrality

SP states that a future history of a general type is metaphysically and technically possible. It does not imply that the posthuman would *improve on* the human or MOSH state or that there would be a commonly accessible perspective from which to evaluate human and posthuman lives. Posthumans may, as Vinge writes, be "simply too different to fit into the classical frame of good and evil" (Vinge 1993).

It could be objected that the value-neutralization of the historical successor relation in the SP schema is excessively cautious and loses traction on what distinguishes humans from their hypothetical posthuman descendants: namely, that posthumans would be in some sense "better" by virtue of having greater capacities.

One of the most widely used formulations of the idea of the posthuman – that of transhumanist philosopher Nick Bostrom – is non-neutral. He defines a posthuman as a "being that has at least one posthuman capacity" by which is meant "a central capacity greatly exceeding the maximum attainable by any current human being without recourse to new technological means". Candidates for posthuman capacities include augmented "healthspan", "cognition" or "emotional dispositions" (Bostrom 2009).

While this is not a purely metaphysical conception of the posthuman it is, it might be argued, not so loaded as to beg ethical questions against critics of radical enhancement. As Allen Buchanan points out, "enhancement" is *a restrictedly value-laden notion* insofar as enhancing a capacity implies making it function more effectively but does not imply improving the welfare of its bearer (Buchanan 2009: 350).

Moreover, it could be objected that "alteration" is so neutral that a technical process could count as posthuman engendering if it resulted in wide descendants of humans with capacities far below that of normal humans (I address this point in Section 5 below).

However, it is easy to see that the value-ladenness of "enhancement" is not restricted enough to capture some conceivable paths to posthumanity. To be sure, posthumans

might result from a progressive enhancement of cognitive powers – much as in Vinge's recursive improvement scenario. Alternatively, our posthuman descendants might have capacities we have no concepts for while lacking some capacities that we can conceive of.

In a forthcoming article I consider the possibility that shared "non-symbolic workspaces" - which support a very rich but non-linguistic form of thinking - might render human natural language unnecessary and thus eliminate the cultural preconditions for our capacity to frame mental states with contents expressible as declarative sentences (Philosophers call such states "propositional attitudes" – e.g. the *belief* that Snow is White or the *desire* to vote for Obama in next presidential election). If propositional attitude psychology collectively distinguishes humans from non-humans, users of non-symbolic workspaces might acquire a non-propositional psychology and thus cease to be human (As I show in section 4 being "human distinguishing" in this manner does not have to entail being part of a human essence).

It is not clear that process leading to this relatively radical cognitive alteration would constitute an augmentation history in the usual sense – since according to my scenario it could involve the loss of one central capacity (the capacity to have and express propositional attitudes) and the acquisition of an entirely new one. Yet it is arguable that it could engender beings so different from us in cognitive structure that they would qualify as posthuman according to SP (See Section 5).

The Borg from the TV series *Star Trek* are a more popular variation on the theme of the "value-equivocal" posthuman. While the Borg seem like a conceivable kind of posthuman life, they result from the inhibition of the kind of cognitive and affective capacities whose flowering Bostrom treats as constitutive of the posthuman. The Borg-Collective, it is implied, possesses great cognitive powers and considerable technical prowess. However, the Collective's powers emerge from the interactions of highly networked "drones", each of whom has had its personal capacities for reflection and agency suppressed.

3. Wide Descent

As advertised earlier, SP uses a notion of *wide descent* to understand our relationship to prospective posthumans.

I will elaborate the distinction between wide descent and narrow descent below in term of a distinction between a *narrow* conception of the human qua species and a *wide* conception of the human. Whereas Narrow Humanity can be identified, if we wish, with the biological species *Homo sapiens*, Wide Humanity is a technogenetic construction or "assemblage" with both narrowly human and narrowly non-human parts.

There are two principle justifications for introducing wide descent and the correlative notion of *Wide Humanity*:

3.1 The appropriate concept of descent for SP is not biological.

Exclusive consideration of biological descendants of humanity as candidates for posthumanity would be excessively restrictive. Future extensions of NBIC technologies may involve discrete bio-technical modifications of the reproductive process such as human cloning, the introduction of transgenic or artificial genetic material or very exotic processes such as personality uploading or "mind-cloning". Thus entities warranting our concern with

the posthuman could emerge via modified biological descent, recursive extension of AI technologies (involving human and/or non-human designers), quasi-biological descent from synthetic organisms, a convergence of the above, or via some technogenetic process yet to be envisaged.

It follows that when considering the lives of hypothetical posthuman descendants we must understood "descent" as relationship that is *technically mediated to an arbitrary degree*.

3.2 "Humanity" is already the product of a technogenetic process.

A plausible analogy for the emergence of posthumans, as Vinge observes, is the evolutionary process that differentiated humans from non-human primates. But there are grounds for holding that the process of becoming human (hominization) has been mediated by human cultural and technological activity. One widely employed way of conceiving hominization is in terms of cultural niche construction. Niche-construction occurs where members of a biological population actively alter their environment in a way that alters the selection pressures upon it. For example, it has been argued that that the invention of dairy farming technology (around 10,000 BC) created an environment selective for genes that confer adult lactose tolerance. Thus the inventors of animal husbandry unwittingly reconfigured the bodies of their descendants to survive in colder climes (Laland, Odling-Smee and Feldman 2000; Buchanan 2011: 204). The anthropologist Terrence Deacon proposes that the emergence of early symbolic practices produced a symbolically structured social environment in which the capacity to acquire competence in complex symbol systems was a clear selective advantage. Thus it is possible that the selection pressures that made humans brains adept at language learning were a consequence of our ancestors' own social activity even as these brains imposed a learnability bottleneck on the cultural evolution of human languages (Deacon 1997, 322-6, 338).

If this model is broadly correct, hominization has involved a confluence of biological, cultural and technological processes. It has produced socio-technical "assemblages" in which humans are coupled with other active components: for example, languages, legal codes, cities, and computer mediated information networks.[1]

Biological humans are currently "obligatory" components of modern technical assemblages. Technical systems like air-carrier groups, cities or financial markets depend on us for their operation and maintenance much as an animal depends on the continued existence of its vital organs. Technological systems are thus intimately coupled with biology and have been over successive technological revolutions.

However, this dependency runs in the other direction: the distinctive social and cognitive accomplishments of biological humans require a technical and cultural infrastructure. Our capacity to perform mathematical operations on arbitrarily large numbers is not due to an

[1] The term "assemblage" is used by the philosopher Manuel Delanda to refer to any *emergent* but *decomposable* whole and belongs to the conceptual armory of the particularist "flat" ontology I will propose for Speculative Posthumanism in Section 4 below. Assemblages are *emergent* wholes in that they exhibit powers and properties not attributable to their parts but which causally depend upon their parts. Assemblages are also *decomposable* insofar as all the relations between their components are "external": each part can be detached from the whole to exist independently (Assemblages are thus opposed to "totalities" in an idealist or holist sense). This is the case even where the part is functionally necessary for the continuation of the whole (DeLanda 2006: 184).

innate number sense but depends on our acquisition of routines like addition or long division and our acculturation into culturally stored numeral systems. Our species-specific language ability puts us in a unique position to apply critical thinking skills to thoughts expressed in public language, to co-ordinate social behavior via state institutions, or record information about complex economic transactions (Clark 2004; 2006). Philosophers such as Donald Davidson and Robert Brandom have gone further, arguing that our capacity to think in and express propositional attitudes depends on our mastery of public language. Davidson argues that the ability to have beliefs (and hence other propositional attitudes such as desires or wishes) requires a grasp of what belief is since to believe is also to understand "the possibility of being mistaken". This in turn requires us to grasp that others might have true or false beliefs about the same topic. Thus no belief can be adopted by someone not already involved in evaluating her own and others' attitudes in common linguistic coin (Davidson 1984).

These considerations lend support to the claim that the emergence of biological humans has been one aspect of the technogenesis of a planet-wide assemblage composed of biological humans locked into networks of increasingly "lively" and "autonomous" technical artifacts (Haraway 1989). It is this wider, interlocking system, and not bare-brained biological humans, that would furnish the conditions for the emergence of posthumans. Were the emergence of posthumans to occur, it would thus be a historical rupture in the development of this extended socio-technical network.

However, while the emergence of posthumans *must* involve the network, the degree to which it would involve modifications of biological humans is conceptually open (as argued above). Posthumans may derive from us by some technical process that mediates biological descent (such as a germ-line cognitive enhancement) or they may be a consequence of largely technological factors.

I shall refer to this wider network as the "Wide Human" (WH). An entity is a *wide human* just so long as it depends for its continued functioning on the Wide Human while contributing to its operations to some degree. Members of the biological species Homo sapiens, on the other hand, are *narrowly* human. Thus, domesticated animals, mobile phones and toothbrushes are wide humans while we obligatory biologicals are both *narrowly* and *widely* human.

Having outlined the patient and the generic process of becoming posthuman, we now state a recursive definition of Wide Human descent:

> *An entity is a wide human descendant if it is the result of a technically mediated process:*
>
> *A) Caused by a part of WH - where the ancestral part may be wholly biological, wholly technological or some combination of the two*
>
> *B) Caused by a wide human descendant.*

A is the "basis clause". It states what belongs to the initial generation of wide human descendants without using the concept of wide descent. B is the recursive part of the definition. Given *any* generation of wide human descendants it specifies a successor generation of wide human descendants.

It is important that this definition does not imply that a wide human descendant *need be human* in either wide or narrow senses. Any part of WH ceases to be widely human if its wide descendants go "feral": acquiring the capacity for independent functioning and replication outside the human network. SP entails that with becoming posthuman this would occur as a result of some history of technical change.

Becoming posthuman would thus be an unprecedented discontinuity in the hominization process. WH has undergone revolutions in the past (like the shift from hunter-gatherer to sedentary modes of life) but no part of it has been *technically altered* so as to function outside of it.

It follows that a wide human descendent is a posthuman if and only if:

I. It has ceased to belong to WH (The Wide Human) as a result of technical alteration.
II. Or is a wide descendant of such a being.

I refer to this claim as the *disconnection thesis*.

4. Disconnection and Anti-Essentialism

My formulation of what it means to cease to be human will seem strange and counter-intuitive to some. We are used to thinking of *being human* not as a part-whole relation (being a part of WH in this case) but as instantiating a human nature or "essence".

An essential property of a kind is a property that no member of that kind can be without. If humans are necessary rational, for example, then it is a necessary truth that if x is human, then x is rational.[2]

To say that a human essence exists is just to say that there is a set of individually necessary conditions for humanity.

Anthropological essentialism (the claim that there is a human essence) implies that the technically mediated loss of even one of these would export the loser from humanity to posthumanity. As metaphysical formula go, this has the immediate appeal of simplicity.

It also provides *a nice clear method for resolving the posthuman impasse*. We can call this the "apophatic method": after the method of apophatic or "negative" theology. Apophatic theologians think that God is so mysterious that we can only describe Him by saying what He is not (Dale 2010). By extension, anthropological essentialism, if true, would allow us to identify each path to posthumanity with the deletion of some component of the human essence. This, in turn, would allow us to adjudicate the value of these paths by considering the ethical implications of each loss of an anthropologically necessary property.

For example, an essentialist may claim on either *a posteriori* or *a priori* grounds that humans are necessarily *moral persons* with capacities for deliberation and autonomous agency. If so, one sure route to posthumanity would be to lose those moral capacities. Put

[2] Another way of putting this is to say that in any possible world that humans exist they are rational. Other properties of humans may be purely "accidental" – e.g. their colour or language. It is not part of the essence of humans that they speak English, for example. Insofar as speaking English is an accidental property of humans, there are possible worlds in which there are humans but no English speakers.

somewhat crudely, we could then know that *some* conceivable posthumans are non-persons. If persons are, as Rawls claims, sources of moral value and non-persons are not then *this* posthuman state involves the loss of unconditional moral status (Rawls 1980). This particular path to posthumanity would, it seems, involve unequivocal loss.

The Disconnection Thesis does not entail the rejection of anthropological essentialism but it renders any reference to essential human characteristics unnecessary. The fact that some wide human descendant no longer belongs to the Wide Human implies nothing about its intrinsic properties or the process that brought about its disconnection. However, we can motivate the disconnection thesis and its mereological (part-whole) conception of wide humanity by arguing against essentialism on general grounds.

The most plausible argument for abandoning anthropological essentialism is naturalistic: essential properties seem to play no role in our best scientific explanations of how the world acquired biological, technical and social structures and entities. At this level, form is not imposed on matter from "above" but emerges via generative mechanisms that depend on the amplification or inhibition of differences between particular entities (For example, natural selection among biological species or competitive learning algorithms in cortical maps). If this picture holds generally, then essentialism provides a misleading picture of reality.

The philosopher Manuel Delanda refers to ontologies that reject a hierarchy between organizing form and a passive nature or "matter" as "flat ontologies". Whereas a hierarchical ontology has categorical entities like essences to organise it, a flat universe is "made exclusively of unique, singular individuals, differing in spatio-temporal scale but not in ontological status" (DeLanda 2002, 58).

The properties and the capacities of these entities are never imposed by transcendent entities but develop out of causal interactions between particulars at various scales. Importantly for the present discussion, *a flat ontology recognizes no primacy of natural over artificial kinds* (Harman 2008).

It is significant that one of Delanda's characterizations of flat ontology occurs during a discussion of the ontological status of biological species in which he sides with philosophers who hold that species are individuals rather than types or universals (DeLanda 2002: 59-60). For example, Ernst Mayr's "biological species concept" (BSC) accounts for species differences among sexually reproducing populations in terms of the reproductive isolation of their members. This restricts gene recombination and thus limits the scope for phenotypic variation resulting from gene flows, further reinforcing discontinuities between conspecifics (Okasha 2002: 200).

Motivated by such anti-essentialist scruples, the bioethicist Nicolas Agar has argued that differences between humans and prospective posthumans can be conceived in terms of membership or non-membership of a reproductively isolated population as conceived by the BSC (Agar 2010: 19). Posthumans would arise where (and only where) radical enhancement created reproductive barriers between the enhanced and the mainstream human population.

Agar's proposal illustrates one variant of the flat ontological approach. However, importing the BSC neat from the science of the evolutionary past is problematic when considering the ontology of technogenetic life forms. Biotechnologies such as the artificial transfer of

genetic material across species boundaries could make the role of natural reproductive boundaries less significant in a posthuman or transhuman dispensation (Buchanan 2009: 352). If these alternative modes of genetic transmission became routinely used alongside regular sex, the homeostatic role of reproductive barriers would be significantly reduced.

While BSC has a clear application to understanding speciation in sexually reproducing life forms, the BSC has no applicability to non-sexually reproducing life forms. Likewise, the distinction between the genetics lab and nature cannot be assumed relevant in a posthuman world where biotechnology or post-biological forms of descent dominate the production of intelligence and the production of order more generally. The flat ontological injunction not to prioritise natural over artificial sources of order provides a more reliable methodological principle than Agar's misguided ethical naturalism.

The distinction between Wide and Narrow Humanity broached earlier in this paper accommodates this possibility by distinguishing between the Narrow Human (which can be understood in terms of the BSC) and the socio-technical assemblage WH which fully expresses human societies, cultures and minds.

WH has the same *ontological status* as species like *Homo sapiens* – both are complex individuals rather than kinds or essences. However, WH is constituted by causal relationships between biological and non-biological parts, such as languages, technologies and institutions. A disconnection event would be liable to involve technological mechanisms without equivalents in the biological world and this should be allowed for in any ontology that supports Speculative Posthumanism.

5. Modes of Disconnection

As mentioned above, Vinge considers the possibility that disconnection between posthumans and humans may occur as a result of differences in the cognitive powers of budding posthumans rendering them incomprehensible and uninterpretable for baseline humans.

For example, he speculates in passing that rich informational connections between posthuman brains (or whatever passes for such) may be incompatible with a phenomenology associated with a biographically persistent subject or self (Vinge 1993).

If non-subjective phenomenology among posthumans is possible then Vinge's concern that such form of existence might not be evaluable according our conceptions of good or evil seem warranted. Human ethical frameworks arguably require that candidates for our moral regard have the capacity to experience pain. Most public ethical frameworks have maximal conditions. For example, liberals valorise the capacity for personal autonomy that allows most humans "to form, to revise, and rationally to pursue a conception of the good" (Rawls 1980: 525).

Autonomy presumably has threshold cognitive and affective preconditions such as the capacity to evaluate actions, beliefs and desires (practical rationality) and a capacity for the emotions, and affiliations informing these evaluations. However, the capacity for practical reason at issue in our conception of autonomy might not be accessible to a being with non-subjective phenomenology. Such an entity could be incapable of experiencing itself as having a life that might go better or worse for it.

We might not be able to coherently imagine what these impersonal phenomenologies are like (e.g. to say of them that they are "impersonal" is not to commit ourselves regarding the kinds of experiences might furnish). This failure may simply reflect the centrality of *human* phenomenological invariants to the ways humans understand the relationship between mind and world rather than any insight into the necessary structure of experience (Metzinger 2004: 213). Thomas Metzinger has argued that our kind of subjectivity comes in a spatio-temporal pocket of an embodied self and a dynamic present whose structures depends on the fact that our sensory receptors and motor effectors are "physically integrated within the body of a single organism". Other kinds of life – e.g. "conscious interstellar gas clouds" or (more saliently) post-human swarm intelligences composed of many mobile processing units - might have experiences of a radically impersonal nature (Metzinger 2004: 161).

Disconnection may take other forms, however. All that is required for a disconnection from the Wide Human recall is that some part of this assemblage becomes capable of going wild and following an independent career. This is not true of current types of artificial intelligence, for example, which need to be built, maintained by narrow humans and powered by other human artefacts. This is why beings that are artificially "downgraded" so that their capacities are less than human are unlikely to generate a disconnection event (See Section 2) – though this possibility cannot be entirely precluded.

A disconnection could ensue, then, wherever prospective posthumans have properties that make their feasible forms of association disjoint from humans/MOSH forms of association.

I suggested in Section 2 that propositional attitude psychology might distinguish humans from non-humans. However, as our excursus into flat ontology shows, the capacity to form propositional attitudes such as the *belief* that Lima is in Peru need not be thought of a component of a human essence but as a filter or "sorting mechanism" which excludes non-humans from human society much as incompatibilities in sexual organs or preferences create reproductive barriers between Mayr-style biological species (Agar 2010: 19-28). Wide successors to humans who acquired a non-propositional format in which to think and communicate might not be able to participate in our society just as our unmodified descendants might not be able to participate in theirs. They would, in this case, "bud off" from the Wide Human, just as a newly isolated species buds off from its predecessors. Such disconnections could happen by degrees and (unlike in a Vingean singularity) relatively slowly relative to the individual lifetimes. There might also be cases where the disconnection remains partial (for example, some non-propositional thinkers might retain a vestigial capacity to converse with humans and MOSH's).

6. Are Disconnections Predictable?

I do not claim that speculations in the previous section reliably predict the nature and dynamics of a disconnection event. For example, we do not know whether greater than human intelligence is possible or whether it can be produced by an "extendible" technological method (Chalmers: 2010).

Nor, at this point, can we claim to have knowledge about the feasibility of the other disconnection events that we have speculated about (e.g. the replacement of propositional attitude psychology with some non-linguistic cognitive format).

These scenarios are merely intended to illustrate the *ontological thesis* that posthuman-human difference would be a discontinuity resulting from parts of the Wide Human becoming so technically altered that they could split off from it. The intrinsic properties exhibited by these entities are left open by the disconnection thesis.

This epistemological caution seems advisable given that the advent of posthumanity is a (currently) hypothetical future event whose nature and precipitating causes are unprecedented *ex hypothesi*. There are many conceivable ways in which such an event might be caused. Even if a Vinge-style singularity is conceivable but not possible some unrelated technology might be a possible precursor to a disconnection. Disconnections are not defined by a particular technical cause (such as mind-uploading) but purely by an abstract relation of wide descent and the property of functional and replicative independence. Disconnection *can be multiply realized by technologies which have little in common other than a) feasibility and b) that disconnection is one of their possible historical effects*. Thus speculating about how currently notional technologies might bring about autonomy for parts of WH affords no substantive information about posthuman lives (even if it may enable a metaphysically and ethically salutary exploration of the scope for posthuman difference).

Assuming that a conceivable technology (For example, controlled nuclear fusion - other than by gravitational confinement in a star) does not violate physical principles the only sure demonstration of feasibility is the production of a working model or prototype. Thus we can have no reliable grounds for holding that conceivable precursors to a disconnection are feasible precursors so long as the relevant technologies are underdeveloped. However, once a feasible precursor has been produced the Wide Human could be poised at the beginning of a disconnection process since the capacity to generate disconnection would be a realized technological power.[3] We may be in a position to know which, if any, of the "usual suspects" (Nanotechnology, Biotechnology, Information Technology, Cognitive Science) might bring about a disconnection only when the potential for disconnection is in prospect.

Thus it is plausible to suppose that any disconnection (however technically realized) will be an instance of what Paul Humphrey terms *diachronic emergence* (Humphrey 2008). A diachronically emergent behaviour or property occurs as a result of a temporally extended process, but cannot be inferred from the initial state of that process. [4] It can only be derived by allowing the process to run its course (Bedau 1997).

If disconnections are diachronically emergent phenomena their morally salient characteristics and effects will not be predictable prior to their occurrence. While this constrains our ability to prognosticate about disconnections, it leaves other aspects of their epistemology quite open. As Humphrey reminds us, diachronic emergence is a one-time event. Once we observe a formerly diachronically emergent event we are in a position to predict tokens of the same type of emergent property from causal antecedents that have been observed to generate it in the past.

[3] Absent defeaters (See Chalmers 2010).

[4] Where the emergent property occurs at the same time as the microstates on which it depends, we have an instance of "synchronic emergence" (Humphrey 2008, 586-7).

Most importantly, that disconnections would be diachronically emergent has no implications for the uninterpretability or "alienness" of posthumans since their nature is left open by the disconnection thesis.

The anti-essentialist flat ontology I have recommended as a basis for the disconnection thesis, gives us grounds to be wary of terms like "uninterpretability" or "alienness". To be sure, posthumans might be strange in ways that we cannot currently imagine. Their human or MOSH contemporaries might struggle unsuccessfully to understand their thoughts or motives. However, the fact that interpretative success is not guaranteed does not entail that relatively alien posthumans would be humanly uninterpretable. There are, after all, many things that we do not understand that we might understand under ideal conditions.

An utterly incomprehensible being ("a radical alien") would not belong to this set. Such a being would be humanly uninterpretable. The inability to understand it would be a necessary or essential part of the human/MOSH cognitive essence. But if, as proposed, we reject taxonomic essences, we must hold that there are no such modal facts of this nature.

It follows that there are no grounds for some holding posthumans to be humanly uninterpretable in principle (i.e. to be radical aliens) since the set of humanly uninterpretable things is not defined. Posthuman thinking may still be so powerful or so strangely formatted that it could defy the interpretative capacities of wide human descendants *not altered to an equivalent degree*. But this would depend on the contingencies of disconnection – which are, as yet, unknown. As pointed out in Section 5, disconnection – like speciation – may come by degrees. If the technology exists to create posthumans, then the same technology might support "interfaces" between human and posthuman beings such as the bi-formatted propositional/non-propositional thinkers mentioned above. Thus where conditions favour it "Posthuman Studies" may graduate from speculative metaphysics to a viable cultural research program.[5]

7. Resolving the Impasse

What are the implications of the disconnection thesis for attempts to negotiate the ethical bind of the posthuman impasse? The impasse is a way of formulating the ethical concern that the posthuman consequences of our own technical activity may be beyond our moral compass. I have conceded that posthumans might be very different from us in diverse ways, but have argued that there is no basis for concluding that posthumans would be beyond evaluation.

As argued in Section 6, we may be in a better position to undertake a value-assessment once a disconnection has occurred. Thus if we have a moral (or any other) interest in *accounting* for posthumans we have an interest in bringing about the circumstances in which accounting can occur. Thus we have an interest in contributing to the emergence of posthumans or becoming posthuman ourselves where this is liable to mitigate the interpretative problems of disconnection.

It could be objected, at this point, that we may also have countervailing reasons for

[5] Vinge alludes to this possibility in his far-future space epic *A Fire Upon the Deep* (Vinge 1992). *In Fire* posthumans so powerful as to be god-like in comparison with the most enhanced transhuman exist on a computationally extreme fringe of space known as "the Transcend" where they are studied by "applied theologians" from observatories on the margins of the Milky Way.

preventing the emergence of posthumans and not becoming posthuman ourselves.

We have acknowledged that some disconnections could be very bad for humans. Since disconnection could go very wrong, it can be objected that the precautionary principle (PP) trumps the accounting principle. Although there is no canonical formulation, all versions of the PP place a greater burden of proof on arguments for an activity alleged to have to potential for causing extensive public or environmental harm than on arguments against it (Cranor 2004; Buchanan: 199-200). In the present context the PP implies that even where the grounds for holding that the effects of disconnection will be harmful are comparatively weak, the onus is on those who seek disconnection to show that it will not go very wrong. However, the diachronically emergent nature of disconnection implies that such a demonstration is not possible prior to a disconnection event. Thus one can use the PP to argue that accounting for disconnection (assessing its ethical implications) is not morally obligatory but morally wrong.

One might conclude at this point that we have substituted one impasse (the conflict between accounting and discounting) for a second: the conflict between the principle of accounting and the PP. However, this will depend on the different attitudes to uncertainty expressed in different versions of the PP. If the principle is so stringent as to forbid technical options whose long-range effects remain uncertain to any degree, then it forbids the development of disconnection-potent technology. However, this would forbid almost any kind of technological decision (including the decision to relinquish a technology).[6] Thus a maximally stringent PP is self-vitiating (Buchanan 2011: 200-1).

It follows that the PP should require reasonable evidence of possible harm before precautionary action is considered. A selective precautionary approach to the possibility of disconnection would require that suspect activities be "flagged" for the potential to produce bad disconnections (even where this evidence is not authoritative). But if disconnections are diachronically emergent phenomena, the evidence to underwrite flagging will not be available until the process of technical change is poised for disconnection.

To take a historical analogy: the syntax of modern computer programming languages is built on the work on formal languages developed in the Nineteenth Century by carried out by mathematicians and philosophers like Frege and Boole. Lacking any comparable industrial models, it would have been impossible for contemporary technological forecasters to predict the immense global impact of what appeared an utterly rarefied intellectual enquiry. We have no reason to suppose that we are better placed to predict the long-run effects of current scientific work than our Nineteenth Century forebears (if anything the future seems more rather than less uncertain). Thus even if we enjoin selective caution to prevent worst-case outcomes from disconnection-potent technologies, we must still place ourselves in a situation in which such potential can be identified. Thus seeking to contribute to the emergence of posthumans, or to become posthuman ourselves, is compatible with a reasonably constrained PP.

Bibliography

[6] Given our acknowledged dependence on technical systems, the long-run outcomes of relinquishment may be as disastrous as any technological alternative.

Agar, Nicholas. 2010. *Humanity's End: Why we should reject radical enhancement*. Cambridge MA: MIT Press.

Bedau, Mark A. 1997. "Weak Emergence". *Philosophical Perspectives* 11:375-399.

Bostrom, Nick. 2009. "Why I Want to be a Posthuman When I Grow Up". In *Medical Enhancement and Posthumanity*, edited by Bert Gordijn and Ruth Chadwick, 107-137. Dordrecht: Springer

Buchanan, Allen. 2009. "Moral Status and Human Enhancement". *Philosophy and Public Affairs* 37 (4): 346-381.

Buchanan, Allen E. 2011. *Beyond Humanity?: The Ethics of Biomedical Enhancement*. Oxford University Press.

Chalmers, David J. 2010. "The Singularity: A Philosophical Analysis". *Journal of Consciousness Studies*.

Clark, Andy. 2004. *Natural Born-Cyborgs: Minds, Technologies and the Future of Human Intelligence*. Oxford: OUP.

Clark, Andy. 2006. "Language, Embodiment and the Cognitive Niche". *Trends in Cognitive Science* 10(8): 370-374.

Cranor, Carl F. 2004. "Toward Understanding Aspects of the Precautionary Principle". *Journal of Medicine and Philosophy* 29 (3): 259 – 279.

Dale, Tuggy. 2010. Trinity. *Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed 28 September 2011. http://plato.stanford.edu/archives/fall2009/entries/trinity.

Davidson, Donald. 1984. "Thought and Talk". In *Inquiries into Truth and Interpretation,* 155-170. Oxford: Clarendon Press.

Deacon, Terrence. 1997. *The Symbolic Species: The Co-evolution of Language and the Human Brain*. London: Penguin.

Delanda, Delanda. 2006. *A New Philosophy of Society: Assemblage Theory and Social Complexity*. London: Continuum.

DeLanda, Manuel. 2002. *Intensive Science and Virtual Philosophy*. London: Continuum.

Haraway, Donna. 1989. "A Manifesto for Cyborgs: Science, technology, and Socialist Feminism in the 1980s". In *Coming to Terms,* edited by Elizabeth Weed, 173-204. London: Routledge.

Harman, Graham. 2008. "DeLanda's Ontology: Assemblage and realism". *Continental Philosophy Review* 41 (3): 367-383

Humphreys, Paul. 2008. "Computational and Conceptual Emergence". *Philosophy of Science* 75 (5): 584-594.

Kevin N. Laland, John Odling-Smee, and Marcus W. Feldman (2000). “Niche Construction, Biological Evolution, and Cultural Change”. *Behavioural and Brain Sciences* 23 (1): 131-146.

Metzinger, Thomas. 2004. *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA.: MIT Press 2004.

Okasha, Samir. 2002. “Darwinian Metaphysics: Species and the Question of Essentialism”. *Synthese* 131 (2): 191-213.

Rawls, John. 1980. “Kantian Constructivism in Moral Theory”. *Journal of Philosophy* 77 (9): 515-572.

Stross, Charles. 2006. *Accelerando*. London: Orbit.

Vinge, Vernor. 1992. *A Fire Upon the Deep*. New York: Tor.

Vinge, Vernor. 1993. “The Coming Technological Singularity: How to Survive in the Post-Human Era”, *Vision-21:Interdisciplinary Science and Engineering in the Era of Cyberspace.* Accessed 8 December 2007. http://www.rohan.sdsu.edu/faculty/vinge/misc/singularity.html.