

The Disconnection Thesis

David Roden (The Open University)

Introduction: The Posthuman Impasse

In a 1993 article “The Coming Technological Singularity: How to survive in the posthuman era” the computer scientist Virnor Vinge argued that a technique for creating entities with greater than human intelligence would lead to the end of the human dominion of the planet and the beginning of a posthuman era dominated by intelligences vastly greater than ours.¹

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.² 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.

¹ Virnor, Vinge, “The Coming Technological Singularity: How to Survive in the Post-Human Era”, *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace* (1993), <http://www.rohan.sdsu.edu/faculty/vinge/misc/singularity.html>.

² The form that this hypothetical intelligence-making technology might take 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 unenvisioned process. All that matters, as David Chalmers argues, is that the method is “extendible”: improving it yields corresponding increases in the intelligence produced. Our only current means of producing human-equivalent intelligence is non-extendible. As Chalmers writes: “If we have better sex . . . it does not follow that our babies will be geniuses.” David Chalmers, “The Singularity: A Philosophical Analysis”. <http://consc.net/papers/singularity.pdf>.

The resulting 'posthuman' minds 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 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 technically feasible, then 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 possible – or not imminent – the Singularity Hypothesis (SH) *still* raises a troubling philosophical problem 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 whose mode of existence might be significantly alien or "other" to ours. I refer to this more general thesis as "Speculative Posthumanism".

Now, it seems we can adopt either of two policies towards the posthuman prospect. Firstly, we can *account* for it by assessing the ethical implications of contributing to the creation of posthumans.

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. The differences stressed in

Vinge's essay are cognitive: posthuman minds might be so different from human minds that we could not understand them. There might be other very radical differences. Posthumans might experience the world in ways that render evaluating a posthuman life from our perspective impossible.

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.³

However, if their technological activity or some future iteration of it could be a posthuman difference-maker, humans and MOSH's have a duty to evaluate these differences with a view to maximizing the chances of achieving the good posthuman outcomes or, at least, avoiding the bad ones. 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. For example, Charles Stross' novel *Accelerando* envisages human and MOSH social systems being superseded by Economics 2.0 – a set of autonomous "resource allocation algorithms" so virulently efficient that conscious life is too slow to keep up and is finally extirpated from the Inner Solar System in the interests of maximising production.⁴

But while this post-singularity scenario is depicted as comically dreadful in Stross' novel, it just does not follow that a posthuman world lacking humans would be dreadful *per se*. If posthumans were as

³ Nicholas Agar, *Humanities End: Why We Should Reject Radical Enhancement* (Cambridge Mass.: MIT 2010), 41-2.

⁴ Charles Stross, *Accelerando* (London: Orbit 2006), 177.

unlike humans as humans are unlike non-human primates, a fair evaluation of their kind of life might be beyond us.

Thus *accounting* for our contribution to making posthumans seems obligatory, if ultimately impossible, while discounting it is irresponsible. We can call this double bind: “the posthuman impasse”.

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

One might try to circumvent the impasse by casting doubt on Speculative Posthumanism. But this is a much weaker claim than SH and thus far more plausible. Vinge’s essay specifies one type of mechanism for generating posthumans. But there might be posthuman difference-makers that do not involve intelligence amplification or 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. So comparable novelty resulting from some combination of biological or technogenetic changes cannot be precluded at some point in the near or distant 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 the metaphysical commitments of Speculative Posthumanism more fully. 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 metaphysical and epistemological account of posthumanity.

Firstly, the SP schema defines posthumanity as the result of a process of technical *alteration*. 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 1.

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 2.

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.

1. 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 exist an accessible common 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”.⁵

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”.⁶

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 the philosophical 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

⁵ Vinge, “The Coming Technological Singularity”.

⁶ Nick Bostrom, “Why I Want to be a Posthuman When I Grow Up”, in *Medical Enhancement and Posthumanity*, ed. Bert Gordijn and Ruth Chadwick (Dordrecht: Springer, 2009), pp. 107-137.

bearer.⁷

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 4 below).

However, it is easy to see that the value-ladenness of “enhancement” is not restricted enough to capture some plausible paths through posthuman possibility-space. To be sure, posthumans might conceivably 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

⁷ Allen, Buchanan, “Moral Status and Human Enhancement,” *Philosophy and Public Affairs* 37, no. 4 (2009): 350, n6.

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 4).

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 just the high-level cognitive and affective capacities that 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.

2. Wide Descent

Such possibilities are thrown into greater relief if we count prospective artificial intelligences or synthetic life forms among our potential successors. 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 biological conception of the human qua species and a wide conception of the human as a social-technical assemblage that includes narrow humans as functionally necessary components. Whereas Narrow Humanity can be identified, if we wish, with the biological species *Homo sapiens*, Wide Humanity is a technogenetic construction 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*:

I) 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 ethical 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 understand “descent” as relationship technically mediated to an arbitrary degree.

II) Humanity is already the technogenetic result of a technogenetic process.

One analogy for the emergence of posthumans, as Virnor Vinge observes, is the evolutionary process that differentiated humans from non-human primates. But there are grounds for holding that this process involved the co-evolution of biological narrow humans, cultural entities such as languages, as well as techniques.⁸ This picture can be integrated into philosophical position known as active-externalism for which the distinction between bodily and extra-bodily processes is irrelevant when identifying

⁸ Terrence Deacon, *The Symbolic Species: The Co-evolution of Language and the Human Brain* (London: Penguin 1997).

cognitive processes such as thinking, memory or imagination. Active externalists like Andy Clark and David Chalmers argue from a principle of “parity” between processes that go on in the head and any *functionally equivalent* process in the world beyond the skin sac. The parity principle implies that mental processes need not occur only in biological nervous systems but in the environments and tools of embodied thinkers. Given parity, spoken language, written documents, culturally embedded crafts, or electronic information systems can all count as vehicles of human mental processes where they make a cognitive contribution to them.⁹ For example, if, as some argue, language is necessary to propositional thought and the latter is humanly distinctive, then human natural language is an obligatory part of human forms of life (See Section 1 and Section 4). Thus just as humans are (for the moment at least) necessary for the continuation of human socio-technical systems, so these systems may be viewed as external preconditions for human life.

Considerations I and II provide grounds for believing that the posthuman would be some kind of *discontinuity* in the technogenesis of humanity and the commencement of some novel technogenesis. II supports the claim that the process has not been exclusively biological thus far, while I) supports the claim that biological processes could be technically mediated or superseded in the emergence of the posthuman.

If so, the relevant *patient* in the process of posthuman emergence would be a system with biological

⁹ Andy Clark and David Chalmers, “The Extended Mind”, *Analysis* 58 no. 1 (1998): 7-19; Andy Clark, *Natural Born Cyborgs* (Oxford: Oxford University Press 2003); ‘Material Symbols’, *Philosophical Psychology* 19, no. 3 (2006), 291–307.

and non-biological components. I shall refer to this complex entity as the 'Wide Human' (WH). WH contains many functionally interdependent components: individual narrow humans, languages, technological systems, legal systems, cities, corporations, religions, nations and geographical regions. It is spatially distributed and temporally evolving. Many of these temporal changes have been due to localized changes in its parts that disseminated through its geographical extent, drastically modifying its powers, which always emerge from the ways its parts interact – e.g. in the ways in which a technology like writing impacts on human cognitive and social practices.¹⁰

These considerations can now inform a recursive definition of Wide Human descent:

An entity is a wide human descendant if it is the historical consequence of a replicative or productive process that is technologically mediated to an arbitrary degree:

A) Occurring to a wide human descendant (recursive part).

B) Occurring to a part of WH (where the ancestral part may be wholly biological, wholly technological or some combination of the two).

3. The Disconnection Thesis

The arguments for value neutrality and wide descent will have hopefully clarified the formulation of SP. They also clarify the sense of “human” and “ceasing to be human” at issue here. If WH is a

¹⁰ For a full exposition of an assemblage-based theory of society see Manuel Delanda, *A New Philosophy of Society: Assemblage Theory and Social Complexity* (London: Continuum 2006).

complex individual and being widely human is a matter of being a functional part of it, then ceasing to be a wide human must consist in ceasing to be a functional part of it.

Ceasing to be human, then, just consists in a sundering or *disconnection* of a part of WH such that this may continue to function independently of it. Let us call this claim the *disconnection thesis*.

This formulation may 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 a complex object) but as instantiating a human nature or “essence”.

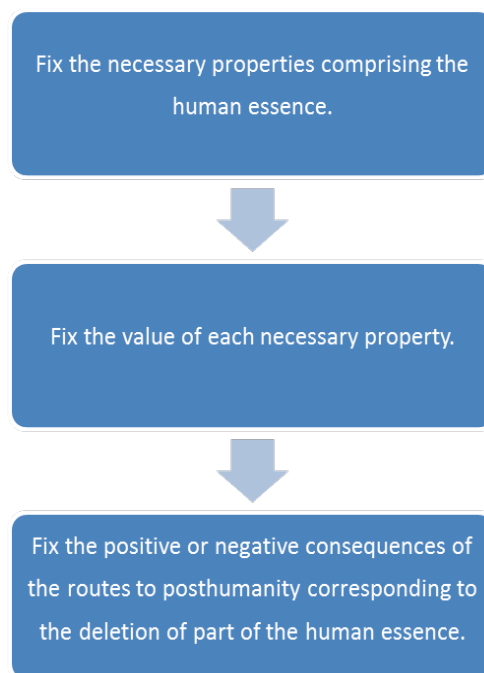
An essential property of a kind is a property that no member of that kind can be without. If humans are necessary rational, for example, then it is a necessary truth that if x is human, then x is rational.¹¹

Anthropological essentialism (the claim that there is a human essence) yields an appealingly simple metaphysics for distinguishing humans and non-humans. To say that a human essence exists is just to say that there is a set of individually necessary conditions for humanity.

Anthropological essentialism implies that the technically mediated loss of even one of these would export the loser from humanity to posthumanity. It also provides *a resolution of the posthuman impasse*. We can call this the “apophatic method”: an allusion to the method of apophatic or “negative” theology. Apophatic theologians think that God is so mysterious that we can only

¹¹ 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. Using the possible worlds formulation of modality, then, we can say that insofar as speaking English is a contingent (or ‘accidental’) property of humans, there are possible worlds in which there are humans but no English speakers.

describe Him by saying what He is not.¹² By extension, anthropological essentialism, if true, would allow us to describe various paths to posthumanity associating each path 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. This can be set out as a three-stage flow chart:



The apophatic route is open to an essentialist whether she adopts the traditional *a priori* view of essences as truths holding in virtue of the meanings of terms or concepts, or the view that the essences of individuals or kinds reflect mind-independent necessary truths that can be discovered *a posteriori*.¹³

¹² Tuggy, Dale, "Trinity", *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2009/entries/trinity/>>.

¹³ For example, prior to modern chemistry and physics it might not have been part of any folk concept of Gold that it is "The element with atomic number 79". Still if it is an *a posteriori* necessary truth that Gold atoms

For example, an essentialist may claim on *a posteriori* or *a priori* grounds that humans are necessary *moral persons* with capacities for deliberation and autonomous agency. If so, one sure route to posthumanity would be to lose those moral capacities. Put somewhat crudely, we could then know that some possible posthumans are non-persons. If persons are, as Rawls claims, self-authenticating sources of moral value and non-persons are not; then entering this posthuman state involves the loss of unconditional moral status.¹⁴ Since no gain in welfare or capacities could compensate for this, the condition becoming posthuman would be an unequivocal evil or loss.

Again, if humans are essentially vicious, then other kinds of posthumanity may be achieved by transcending viciousness (through some biomedical alteration, perhaps). Again, depending on our list of essential properties, the impasse would be resolvable by apophasis.

The Disconnection Thesis does not entail the rejection of anthropological essentialism but it renders anthropological essentialism irrelevant to SP. That a temporal stage m_t of a particular wide lineage L_w ceases to belong to the Wide Human in some future history implies nothing about the properties of the members of m_t other than that they are no longer parts of WH at t . However, we can motivate the disconnection thesis and its mereological (part-whole) conception of wide humanity broached above by arguing against essentialism on general ontological grounds.

Metaphysical nominalists have traditionally motivated the rejection of essences and similar abstract entities on grounds of ontological parsimony. But there are also realist grounds for denying essences

have 79 protons and 79 electrons, having the atomic number 79 is part of the essence of that metal. See Saul Kripke, *Naming and Necessity* (Cambridge: Harvard University Press 1980).

¹⁴ John Rawls, "Kantian Constructivism in Moral Theory", *The Journal of Philosophy* 77 no.9 (1980): 515-572.

a fundamental ontological role in accounting for the way the world is structured taxonomically – e.g. in terms of biological species or chemical kinds. Essentialism supposes that differences in material nature are formally structured in a way that could be represented in an ideally true linguistic representation of it. Thus for an essentialist version of realism, things are sorted into kinds in virtue of the essential properties of those kinds.

Anti-essentialist realists like Manuel Delanda argue that essentialism replicates an ancient “hylomorphic” conception of the world according to which each thing consists of a compound of unformed matter and intelligible form (e.g. according to Aristotle, the “soul” is the form, or mode of organization, of a living entity). However, we now know that form is not imposed on matter from without but emerges via generative mechanisms in which essences play no role (For example, natural selection among biological species or competitive learning algorithms in cortical maps). If this picture holds generally, then essentialism provides a misleading account of reality.

The limitations of hylomorphism are particularly apparent in post-Darwinian biology. Many philosophers of biology hold that the interpretation of biological taxa most consonant with Darwinian evolution is that they are individual populations differentiated from other populations by reproductive, ecological and symbiotic relationships rather than a common essence. A population (not an essence or kind) can be differentiated and variable and, in consequence, can undergo the self-organizing feedback processes required by the theory of natural selection.

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 or concepts to organise it, a flat universe is “made exclusively of unique, singular individuals,

differing in spatio-temporal scale but not in ontological status”.¹⁵ 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.¹⁶

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.¹⁷ 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.¹⁸

These populations are identified not by their exemplification of a common essence but by homeostatic mechanisms which maintain the cohesion of gene flows, generating cohesive patterns of co-variation in phenotypic traits that we pick up on in our folk-categorizations of felines, canines, tulips, etc. Population thinking allows us to explain the processes of self-organization that result in functionally complex assemblages purely in terms of interactions within differentiated populations.¹⁹

¹⁵ Manuel DeLanda, *Intensive Science and Virtual Philosophy* (London: Continuum 2009), 58.

¹⁶ Graham Harman, “DeLanda's ontology: Assemblage and realism,” *Continental Philosophy Review* 41 no.3 2008, 372.

¹⁷ Delanda, *Intensive Science and Virtual Philosophy*, 60.

¹⁸ Samir Okasha, “Darwinian metaphysics: Species and the question of essentialism,” *Synthese* 131 no.2 2002: 200.

¹⁹ Delanda, *Intensive Science and Virtual Philosophy*, 59.

Motivated by such anti-essentialist scruples, Nicolas Agar argues 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.²⁰ 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. Allen Buchanan points out that biotechnologies such as cell nuclear replacement and 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.²¹ If these alternative modes of genetic transmission became routinely used alongside regular sex among humans/non-humans/MOSH's then the homeostatic role of reproductive barriers would be significantly lessened. Would this mean that the BSC would have no application to the status of posthumans? Agar is understandably concerned to rebut this implication:

This argument, however, overstates the problem for the biological species concept. Facts about what might or might not happen in a laboratory are actually not directly relevant to decisions about species boundaries. We assign individual organisms to species not on the basis of what happens in human or posthuman laboratories, but according to what occurs in nature. Humans and chimpanzees cannot interbreed in nature even if they might do so in a specific laboratory setting.²²

²⁰ Agar, *Humanity's End*, 19.

²¹ Buchanan, "Moral Status and Human Enhancement," 350, n10.

²² Agar, *Humanity's End*, 22.

However, 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, it would have little explanatory value in a world where reproductive boundaries were widely breached by technological mediators. The distinction between the artificial conditions that apply in a genetics lab and the natural conditions in which species evolve over biological time cannot be assumed relevant in a posthuman world where advanced biotechnology or post-biological forms of descent dominate the production of intelligence, and the production order more generally. Where artificial generative processes are liable to become dominant ways of generating order in the world (as might be the case prior to during a posthuman dispensation) there is clearly little point in carving reality according to generative process which have ceased to form it. Here the flat ontological injunction not to prioritise natural over artificial sources of order provides a more reliable methodological principle than Agar's ethical naturalism.

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

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 where the non-biological components, such as languages, technologies and institutions, have been characterized by much faster rates of change than the biological parts. 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.

4. 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, Vinge speculates 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.²³

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 phenomenologies similar to normal human phenomenologies. For example, our moral regard for animals is conditional on their presumptive capacity for experiencing pains and pleasure. 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".²⁴

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 cognitive system with non-subjective

²³ Vinge, "The Coming Technological Singularity".

²⁴ Rawls, "Kantian Constructivism", 525.

phenomenology. It could be incapable of experiencing itself as having a unique biography that might go better or worse for it. It might have belief-like self-representations of its physical substrate or its various internal processes (like my current thought about a region on my big toe) but could not experience humiliation, resentment or satisfaction because these attitudes require a rich apperceptive experience of persistent subjectivity.

But is this phenomenological speculation plausible?

Following Kant, some philosophers in the transcendental tradition have argued that persistent subjectivity is a condition of knowledge and rationality. One might argue along these lines that these hypothetical phenomenologies would be incompatible with general intelligence or rationality. Thus if posthumans are assumed to be at least as generally intelligent or rational as we, it would follow that their phenomenologies could not be so alien.

Given its commitment to value neutrality, the comparability of humans and posthumans is not assumed by Speculative Posthumanism. Nonetheless, the possibility that posthumans may be both alien and possess greater-than-human capacities to transform our shared environments is a legitimate source of concern. Is it, then, coherent to suppose that there might be beings with enormous general intelligence but radically non-human phenomenology?

Although I cannot present the argument in detail, here, I think there are grounds for concluding that we simply do not know enough about phenomenology to preclude such entities. Some philosophers confidently assert that we have a “direct” or intuitive grasp of what states of consciousness are possible or impossible but the epistemic basis for such insights is profoundly questionable.

For example, the author of a recent primer in the philosophy of mind claims that having judgments

of perception without the relevant perceptions is impossible: "[If] I am aware that I am having visual experiences, say, then surely I must actually be having visual experiences".²⁵ We can hallucinate how things are, but we cannot be fooled into believing that the world *seems* a certain way if it does not seem that way.

To be sure, it is hard to *imagine* how our subjective judgments regarding how things seem could be false, but, as Daniel Dennett has observed, it is a liability of the philosophical profession to mistake failure of imagination for insight into necessity.²⁶ The impossibility claim regarding subjective hallucination appears refuted by conditions such as Antons's syndrome where sufferers report visual experiences that they are not capable of having due to damage to their visual cortex.²⁷

If, as some philosophers claim, we can account for the characteristic structure of our phenomenology in terms of the functional roles of sensorimotor systems or forms of mental simulation then it is legitimate to speculate on the scope for other role-fillers. Maybe all intelligences need some *Dasein* (to creatively abuse Heidegger's term for our specifically human way of being temporally present in the world). Thomas Metzinger argues that our kind of *Dasein* 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 *Dasein's* of a radically impersonal nature.²⁸

²⁵ Sean Crawford, *Aspects of Mind*. (The Open University: 2005), 43.

²⁶ Daniel C. Dennett, *Consciousness Explained* (London: Penguin 1991), 401.

²⁷ Thomas Metzinger, *Being No One: The Self-Model Theory of Subjectivity*, (Cambridge, MA.: MIT Press 2004, 243.

²⁸ Metzinger, *Being No One*, 161.

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 specific modes of *Dasein* they 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.²⁹

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 assuming 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 1) – 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 1 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 (Some humans – neonates and impaired adults - may not be capable of having mental states with propositional content, but are no less human for all that).

Rather than as a constituent of a human nature or essence, propositional thinking can be understood 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

²⁹ Metzinger, *Being No One*, 213.

Mayr-style biological species.³⁰ 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).

It is arguable that actual disconnections could present hard cases where it is debatable whether the putative posthuman population has *really* sundered from the human much as ordinary judgements of identity and difference between institutions, countries, organisms, species, etc. can encounter “vague” or borderline cases. However, as long as it is reasonable to say that individuals can be distinct (in spite of the acknowledged hard cases) it remains reasonable to employ disconnection as a framework for thinking about posthuman-human difference.

5. Are Disconnections Predictable?

I do not claim that speculations in the previous section are reliably prognostic. For example, we do not know whether greater than human intelligence is possible, let alone whether it can be produced by an “extendible” technological method.³¹ Nor, at this point, can we claim to have knowledge about the feasibility of the other disconnections events that we have speculated about (e.g. the replacement of propositional attitude psychology with some non-linguistic cognitive format).

³⁰ Agar, *Humanity’s End*, 19-28.

³¹ Chalmers, “The Singularity: A Philosophical Analysis”.

Thus the aforementioned disconnection scenarios are best seen as attempts to consider the scope for possible disconnections. Not as reliable “geographies” of posthuman possibility space.

This epistemological caution seems advisable given that the advent of posthumanity is a (currently) hypothetical future event whose ethical and metaphysical interest derives from its potential for novel forms of life or association whose transformative powers are without empirical precedent (Bottom line: *there has not been, to date, a human-posthuman disconnection!*). Thus it seems that disconnection, at minimum, be an instance of what Paul Humphrey terms *diachronic emergence*.³²

An emergent property of a complex system S is a property of the whole of S (a macrostate) rather than of the parts of S which depends (or “supervenes”) on the “microstates” of S’s parts but which has some “autonomy” with respect to those microstates. The difficulty of reconciling the autonomy of emergent phenomena with their micro-dependence makes the metaphysics of emergence a lively area of philosophical debate.³³ For example, some emergentists claim that emergent properties exercise “downward” causal influence on the microbehaviour of the systems in which they are exhibited. This grade of metaphysical autonomy is usually categorized as “strong emergence”.

In the case of diachronic emergence, however, the autonomy of the emergent phenomenon is purely “inferential”. 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.³⁴ Thus

³² Paul Humphreys, “Computational and Conceptual emergence,” *Philosophy of Science* 75 no. 5 2008: 584-594.

³³ Jagewon Kim, “Emergence: Core Ideas and Issues”, *Synthese* 151 no. 3 2006: 547-559.

³⁴ Humphrey, “Computational and Conceptual Emergence”, 587-8. Where the emergent property occurs at the same time as the microstates on which it depends, we would have an instance of “synchronic emergence”.

treating singularities and other posthuman difference-makers as diachronically emergent implies a lower grade of metaphysical commitment since diachronic emergence is a form of “weak emergence” (Humphrey 2008, 587). Whereas the existence of strongly emergent properties is heavily disputed by many philosophers and scientists, it is widely accepted that the future behaviour of some systems is “computationally incompressible” – i.e. underivable by a process simpler than the process which actualizes that behaviour (Humphrey 2008, 588). Mark Bedau defines a weakly emergent property as follows. Suppose a system S has macrostates constituted wholly by microstates of its parts and that it has a dynamic D which determines the evolution of its microstates:

Macrostate P of S is weakly emergent iff P can be derived from D and S’s external conditions
*but only by simulation.*³⁵

If (as seems plausible) singularities and other posthuman difference makers would exhibit such novel behaviour as to belong to the class of diachronically emergent events, then their salient characteristics would not be predictable short of either waiting for them to occur or running a simulation of equivalent complexity and transformative power. For example, any revealing simulation of the impact of a synthetic intelligence with human-equivalent intelligence or greater would be liable to exhibit similar transformative power.³⁶

³⁵ Mark A. Bedau, Weak emergence, *Philosophical Perspectives* 11 1997: 377-8. My emphasis.

³⁶ By contrast, simulations of some aspects of disconnection events may be relatively simple computationally but uninformative about matters of real concern. Suppose our hypothetical posthuman swarm intelligence were a virus that parasites the energy and processing capacity of the human nervous system and is spread from host to host by induced sneezing. Its dissemination would presumably be predicted with epidemiological models that work for diseases like influenza but such models of disease spread would tell us little about its effects on the world or upon its human hosts.

So a disconnection event is liable to be its best and only adequate simulation.

However, while this constrains our ability to prognosticate about disconnections, it leaves other aspects of their epistemology 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 typically generate it.

Moreover, that disconnections would be diachronically emergent has no implications whatsoever for the uninterpretability or “alienness” of posthumans. 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 concretely strange in ways that we can barely imagine and their humans or MOSH contemporaries (if such there be) might be hard put to understand their thoughts or motives. However, it does not follow that relatively alien posthumans would be “uninterpretable” for humans/MOSH’s in principle. There are many things that we do not understand that we might understand after inordinate effort or under ideal conditions. A necessarily incomprehensible being (“a radical alien”) would not belong to this set. Such a being would be humanly uninterpretable. That is: 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 facts.

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 historical circumstances of humans and posthumans alike. As pointed out in Section 4, 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 conditions are relatively favourable “Posthuman Studies” may graduate from speculative metaphysics to a viable cultural research program: though whether this proves fruitful or “degenerating” would depend entirely on the nature and the dynamics of the disconnection.³⁷

Conclusion: 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. However, while we have conceded that posthumans might be fundamentally different from us in many ways, we have no basis for concluding that posthumans would so radically alien as to be beyond evaluation. There are too many conceivable paths to posthumanity to support such a judgement. However, if, as argued in Section 5, these are all diachronically emergent phenomena then humans or MOSH’s will not be in a position to evaluate the strangeness or value of a posthuman state prior to the advent of posthumans.

However, as argued in Section 5, we may be in a better position to undertake such an assessment once disconnection has occurred and posthumans exist. Thus if we have a moral (or other) interest

³⁷ Vinge alludes to this possibility in his far-future space epic *A Fire Upon the Deep*, (Tor 1993). In *Fire* posthumans so powerful as to be “weakly god-like” compared 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.

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 that we may also have countervailing reasons for *preventing* the emergence of posthumans and not becoming posthuman ourselves. We have acknowledged that some conceivable disconnections might be terminally bleak. Since disconnection could go very wrong, it can be objected that the precautionary principle trumps the principle of accounting for the long-run impact of our technical activity. If it is possible that this impact might be catastrophic, then, so the argument goes, the onus is on those who seek to bring about a disconnection to demonstrate that its consequences would be virtuous rather than otherwise. Since I have argued that this is not possible, accounting, it could be claimed, is not morally obligatory, but 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 principle of precaution. However, if (as seems reasonable) the precautionary principle cannot be applied to all technical activity, a precautionary approach to the possibility of disconnection requires that suspect activities must be “flagged” for any disconnection-inducing potential.³⁸ However, if disconnections are diachronically emergent phenomena, such potential may only be demonstrable when the process of technical change is relatively advanced. To take an analogy: the syntax of modern computer programming languages is built on the work on formal languages developed in

³⁸ One could argue that total technological relinquishment is always an option. However, if human beings are technical beings (See Section 2) then this is not an option for us. Moreover, given our current dependence on technical systems, the long-run outcomes of relinquishment may be both as conceivably disastrous as any technological alternative. The principle of precaution should counsel against it.

the 19th 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 have predicted the massive global impact of what appeared, at the time, to be 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 19th Century forebears. Thus even if we acknowledge the need for reasonable caution in the development of technologies with the potential to disconnect posthumans from humans, we must still place ourselves in a situation in which that potential can be identified. Thus our interest in contributing to the emergence of posthumans, or becoming posthuman, is compatible with a reasonable interpretation of the precautionary principle.