2 A defence of pre-critical posthumanism

Introduction

As we have seen, critical posthumanists think speculations about such technically engendered posthuman successors as digitally emulated minds, synthetic life forms or robots evince a clumsy naivety. They argue that futurists who worry about roboapocalypses or who dream of becoming an immortal soul engine fail to understand that their fantasies of transcendence or annihilation replicate humanist assumptions about the universal nature of human reason, the dispensability of bodies or the stability of the human essence (§§1.3, 1.4). They fail to grasp that the "human" to which such hypothetical beings are "post" is already a historically variable cultural and technological construction. There can be no posthuman successor species because, as the title of Hayles's *How We Became Posthuman* implies, we have already entered a posthuman dispensation in which the very value and status of the human is put in question by developments in science, political theory and philosophy. If this view is correct, then critical posthumanism is antithetic to SP, and perhaps the only posthumanism worth fighting for.

However, in this chapter I will argue that critiques advanced by theorists like Wolfe and Hayles misdiagnose futurist posthumanism as a technologically intensified version of humanism. This derives, in part, from the basic conflation of SP and transhumanism trailed in the previous chapter. While critical posthumanists make palpable hits on certain technological fantasies of transcendence, SP is not committed to these. It claims that a nonhuman successor to humans could arise in consequence of our technological activity. But it is not committed to the claim that such beings will realize our humanist dreams or apocalyptic nightmares.

Analysing why these arguments fail has the dual benefit of preventing us from being distracted by the anti-humanist hyperbole accruing to theoretical frameworks employed in critical posthumanism – such as deconstruction and cognitive science – but, more importantly, contributes to the development of the rigorous, philosophically self-aware speculative

posthumanism that I hope to develop in this book. For example, it will bring into view the extent to which SP is committed to a minimal, non-transcendental and nonanthropocentric humanism and will help up put bones on its realist commitments (see Chapter 5).

This chapter will consider four "dismissals" of SP that occur within the extant literature of critical posthumanists:

- The anti-humanist argument
- The technogenesis argument
- The materiality argument
- The anti-essentialist argument.

All four, I hope to show, are unsound.

2.1 The anti-humanist argument

Critical posthumanists with a deconstructive orientation often accuse speculative futurists of misconceiving the posthuman as a radical break with, or transformation of, the human condition. On the contrary, Hayles and Haraway argue: we are *already posthuman*, living on after our deeply machinic, inhuman nature has been exposed by theories like deconstruction and cognitive science, and by the practical enmeshing of the human body-subject within ramifying NBIC technologies such as biomedicine and cybernetics:

> When the self is envisioned as grounded in presence, identified with originary guarantees and teleological trajectories, associated with solid foundations and logical coherence, the posthuman is likely to be seen as antihuman because it envisions the conscious mind as a small subsystem running its program of self-construction and self-assurance while remaining ignorant of the actual dynamics of complex systems. But the posthuman does not really mean the end of humanity. It signals instead the end of a certain conception of the human, a conception that may have applied, at best, to that fraction of humanity who had the wealth, power and leisure to conceptualize themselves as autonomous beings exercising their will through individual agency and choice.
>
> (Hayles 1999: 286)

SP depicts humanity as determinably different from its others – such that the posthuman would constitute a radical break with it. If we are already not the humans that we thought we were, the possibility of rupture through the emergence of posthumans is foreclosed.

Apocalypse postponed?

This is too quick. There are, as we have noted, many ways in which humans might be distinguished from nonhumans: for example, as a transcendental subject or as a member of a distinctive biological species. The "human being" at issue in this passage from Hayles's *How We Became Posthuman* is not the transcendental subject of Kant or Husserl or *Homo sapiens* but the autonomous moral subject that can reason about its commitments and its plans for life.

So is Hayles at least right to claim that our conception of autonomy has been complicated to the point at which we should reject the distinction between autonomous human persons and "heteronomous" things, machines, animals? This would be a hit against one important component of SP: the claim that posthumans might be significantly *weird*. Vinge suggests, for example, that posthuman life might be more akin to Lovecraft's amorphous "elder gods" than the merely uncanny Cylons or Replicants (Lovecraft 1999; §4.3). If we are already alienated from the subjects we thought we were, then Vinge and Lovecraft's radical aliens might not be as weird as all that.

I have christened this "the anti-humanist objection to posthumanism" because it draws heavily on arguments by prominent French anti-humanists like Derrida and Deleuze. Derrida – as we have seen – argues that what we take to be a unified subject is a complex field of relations: a generalized writing machine whose thought is articulated by events and structures that it cannot control.

In the posthumanist critiques of Hayles and Haraway this deconstructive picture is beefed up with models and theories drawn from cognitive science, complexity theory and cybernetics.

Let us consider cognitive science first.

Classical and embodied cognition

Hayles's account of the posthuman subject is influenced by "embodied" approaches to cognition that emphasize the dependence of thought on its embodiment and material relationships.

Embodied cognition belongs to a number of revisionary responses to so-called "Classical" cognitive science. Classicism in cognitive science draws heavily on Descartes' internalist picture of minds as abstract representational systems for which the body and its environment are mere input devices. All the interesting mental stuff goes on after the body's interaction with its world has caused ideas or concepts to appear in the mind (Wheeler 2005; Samuels 2010; Fodor 1980).

Classicism pictures brains as akin to stored program computers containing discrete symbolic representations of their environments. Mental

processes consist in the transformation of these elements according to computational rules or "algorithms". The mental symbols are defined purely by their characteristic shape or structure (their *syntax*) and not by their *semantic properties* (their meaning or content). Formal principles of reasoning are analogous to the rules of game that determine how one position in the game can be succeeded by a later position. They determine how one sentence in the calculus can be derived from another by virtue of its syntactic structure *irrespective of what the symbols mean*.

As an example of how such structure-sensitive rules might yield something like cognition, suppose we have a program with a branching 'If … Else' rule such as:

If p = "horns", then add x to list [Goats],
Else add x to list [Sheep]

Intuitively, this is a *very* simple categorization rule. It specifies a file location represented by the variable letter "p" and it instructs the computer to add the information contained in another file location "x" to one of two lists [Goats] and [Sheep], depending on whether the data in p corresponds to the character string "horns". The choice of branch comes down to the "shape" of the data stored in particular physical locations on a system that implements this program.

While the example is super simple, the moral is general. *Computational rules determine actions whose execution depends on the local properties of symbols to which they are applied*. The rules themselves are blind to the meaning of those representations.

If reasoning in human and nonhuman animals proceeds in this way, it ought to be possible to explain understanding and meaning by positing "dumb" computational engines whose components lack the florid mental powers they render possible.

The artificial intelligence pioneers Herbert Simon and Allan Newell argued, accordingly, that treating minds as "physical symbol systems" whose state transitions are governed by syntax-sensitive rules is a scientifically fruitful way of explaining how humans and animals think and offers clues for producing artificial intelligences implemented on similar lines (Newell & Simon 1976).

Embodied cognitive science, by contrast, draws inspiration from computational prowess exhibited in biological systems which exhibit no symbolization. Its proponents argue that the preconditions of intelligence can emerge from local interactions between relatively dumb agents (like ants or neurons) and their environments without a planner or "thinking subject" to choreograph their activities; and without the need for computational rules like the branching if/else statement. Swarm intelligence

is one example of such emergent computation. It is exhibited where a population of interacting agents – for example, ants, robots or software entities – displays a problem-solving capacity that is not possible for any individual within the population working in isolation. For example, Deneubourg *et al*. (1990) showed that colonies of Argentine ants (*Iridomyrmex humilis*) were able to discriminate a nearer food source from a more distant one by utilizing a simple positive feedback relation between pheromones deposited by ants and the hard-wired tendency of ants to head towards the greatest concentrations of pheromone. Since pheromones evaporate with time, the trails left by ants returning from nearer food sources tend to have greater concentrations, thus *recruiting* more ants for foraging and catalyzing recruitment by further biasing the density of the pheromone trail. Here the problem-solving power of the hive "superorganism" *emerges* from the positive feedback between ant recruitment and pheromone signals. This is one of many cases in which component interactions induce *self-organizing behaviour* in a complex system without the need for a central controller or queenly "hive mind" to choreograph their efforts (see also discussion of neural networks in §4.1).

The classical, symbol-driven approach has also been subject to an extensive critique by embodied theorists like Wheeler, Clark, Rodney Brooks and Susan Hurley. This is motivated by some apparent difficulties facing the Classical account.

One way of understanding these is by distinguishing between a process being *computable* and being *tractably computable*. In the early part of the twentieth century it was shown by Alan Turing and others that not all well-defined mathematical functions are computable. There are functions whose values for particular inputs cannot be calculated by following formal rules for manipulating logical symbols. Turing made the conceptual leap from formal logic to computation by showing that inferences in formal systems can be implemented on mathematically defined symbol-manipulating devices (now called "Turing Machines").[1] A Turing machine manipulates symbols according to a machine table (i.e. its program) specifying how it should behave when reading a particular symbol at a particular memory location on its "tape", when in a particular state. The scanning and machine behaviours are extremely simple operations that can be accomplished by any physical machine which responds differentially to its internal states. Moreover, since their rules are purely syntax-sensitive we need not ascribe an understanding of meaning to any part of the computer responsible for these operations (Petzold 2008). Many writers use the idea of a Universal Turing Machine – an abstract computer that can read any "program" on its tape and compute the result – to capture the intuitive idea of a mechanical computation. According to this (contested) view, any computation that can be undertaken by a physical

information processor can be represented by a program-controlled Universal Turing Machine (Copeland 2000).

However, a Turing-computable function may not be *tractably computable* within real-world constraints on time, memory and energy. Tractability matters for humans, animals or replicant fugitives who must respond fluidly to whatever the world throws at them. According to classical computational theory of mind, they achieve this via a four-stage process dubbed the *Sense-Model-Plan-Act* cycle (SMPA) by Rodney Brooks. The first stage of the cycle is to acquire sensory information from input devices (eyes, cameras, nose, whiskers, etc.). The second is to construct symbolically represented beliefs about the world from the sensory information. The third (Plan) is to infer a series of actions by applying general structure–sensitive rules to the beliefs formed in the second stage. The fourth stage is to generate those actions by transforming the plan into a structured series of movements (Brooks 1991: 140; Wheeler 2005: 67).

Critics of the symbol system approach to understanding cognition, like Wheeler and Hubert Dreyfus, have argued that Steps 2 and 3 are particularly problematic in any complex environment. In particular, both tasks are subject to what has come to be known as "frame problems". Frame problems concern how a cognitive system distinguishes relevant from irrelevant information. Humans and higher nonhuman animals regularly make skillful and occasionally very fast inferences about the state of their world. Here are some examples – the last two due to Churchland (2012):

- There are voices coming from the empty basement – the DVD has come off pause!
- Smoke is coming out of the kitchen – the toast is burning!
- *Artificial* selection of horses, pigeons, pigs, etc. can produce new varieties of creature – evolution is *natural* selection!

Frame problems pose a challenge to classicism because it implies that a sophisticated cognitive system has myriads of belief-like entities internally represented as discrete physical states (like "inner sentences" or file locations in Random Access Memory). Even Fodor, the arch-classicist, concedes that these feats of fluid inference are hard to explain because it requires our brains to put a "frame" around the representations relevant to making the inference – information about the Highway Code or the diameter of the Sun probably won't be relevant to figuring out that burning toast is causing the smoke in the kitchen. Relevance seems to be a holistic property – beliefs are relevant given a context, given our values and in virtue of *relations to lots of other beliefs*.

But which ones? How do our brains know where to kink the frame without first making a costly, unbounded search through all our beliefs, inspecting each for its relevance to the problem?

Suppose that a system must evaluate the consequences of some plan for its world-representation. Unless it first isolates a small subset of relevant beliefs, it will need to iterate through a World Stack, a list containing all its beliefs arranged in no particular order to pick out the ones that will be relevant to the task at hand (the order would be indifferent since prior to evaluation no relevance score can be assigned to them). If its World Stack is very large, the system is liable to have to plough through many entirely irrelevant beliefs, deducing whether it has any implications for its plan. With a very big World Stack, this process may turn out to be computationally intractable. After all, unless there is a time limit on updating, the system may need to review beliefs it has updated earlier in the light of updates of subsequent beliefs in the stack. Maybe the system will clunk away evaluating many irrelevant beliefs before it gets to the ones it needs to reassess in the light of its plan (Fodor 1983: 112–13; Wheeler 2005: 178–82).

As Terrence Horgan and John Tienson point out (using Fodor's analysis) this problem is compounded by the fact that a sophisticated, rational system will also need to be sensitive to non-local properties of a world-representation like simplicity and conservatism (Horgan & Tienson 1994: 314). Horgan and Tienson think that it is quite likely that holistic properties like relevance and non-local properties like simplicity cannot be captured in computationally tractable algorithms.

For critics of classicism, the frame problems are symptomatic of the need for a different approach to understanding the mental – though, even among critics of the physical symbol approach, there is little consensus on what this ought to be (see Churchland 2012; Horgan & Tienson 1994).

Wheeler argues that the frame problem for mental representation arises because the representations in question are presumed *disembedded* from their environmental contexts:

> In typical cases of perceptually guided intelligent action, the environment is not more than i) a furnisher of problems for the agent to solve, ii) a source of informational inputs to the mind (via sensing), and, most distinctively, iii) a kind of stage on which sequences of preplanned actions (outputs of the faculty of reason) are simply executed.
>
> (Wheeler 2005: 45)

This means that the environment only supplies raw material for the construction and updating of inner representations and the occasions for

action but *plays no role in mental processing*.[2] How does disembedding mental representation contribute to problems of relevance?

Well, here Wheeler turns to Martin Heidegger's phenomenological critique of rational psychology in his *Being and Time*.

If thought is the manipulation of mental representations according to rules sensitive purely to their physical structure, then context can only figure in thought by being *explicitly represented* by the symbols which refer to objective features of that context like shapes or motions. Otherwise structure–sensitive rules can take no account of it.

Heidegger argues that this Cartesian model of the mind as a representational system over-intellectualizes human agency. While humans can create explicit representations of objects, their everyday access to the world is that of skillful, engaged coping. In skillful coping we rarely represent objects explicitly, according to Heidegger. More commonly, we are aware of objects in terms of their significance for current tasks (as *zuhanden* or ready-to-hand) while we are aware of our environment primarily as a set of potentials for action (sometimes referred to as "affordances") not as represented bundles of properties and relationships (this idea will be important when we consider the worldly background of interpretative understanding in §3.7).

According to Heidegger, this phenomenology of everyday agency belies the classical picture of a rational subject representing mind-independent properties of an external world in sentences and concepts (Heidegger 1962; Dreyfus 1990).

Heidegger's phenomenologically based insight into the structure of everyday coping can inform an embodied cognitive science by shifting efforts away from representing knowledge in terms of inner symbols and rules and towards the kind of skillful, flexible coping activity that biological organisms exhibit on their home turf. In AI and cognitive science, for example, this approach is evident in behaviour-based robotics systems which build sparse and temporary representations of their world by sensing the current context and activity of the robot. According to Wheeler, such "action-oriented representations" *build in* value and contextual relevance to the system *because of their dependence on the situation of a robot or organism*. Here the world is "encoded in terms of possibilities for action" much as Heidegger inferred from his phenomenological account of human agency (Wheeler 2005: 197). Relevance does not have to be deduced by computing the outcomes of lots of individual facts (spawning the frame problem) since action-oriented representations are inherently value-laden and relevant.

The mind portrayed by the embodied approach is not, then, the Cartesian-internalist mind standing apart from its world, but an externalizable pattern of bodily interactions, a patterning which, as in ant

superorganisms, can emerge from asynchronous interactions between dumb components. According to this "active-externalist" picture, Hayles argues, *there is no classically self-present human subjectivity for the posthuman to transcend.* Mental powers of deliberation, inference, consciousness, etc. are *already* distributed between biological neural networks, actively sensing bodies and artefacts (Hayles 1999: 239, 286). *Pace* Descartes, humans are not self-transparent subjects but beings that appear remarkably inept at understanding their nature. Never mind posthumans – humans are already weird amalgams of machines. We just don't know it yet.[3]

Deconstruction

What of the deconstructive attack on the autonomous subject that Hayles takes to complement that of cognitive science? As we saw in §1.4, Derrida argues that subjectivity depends on generalized writing or general textuality, where the notion of a "general text" refers to a highly abstract set of conditions for the production of "sense" or "meaning" which any signifying item must satisfy. For example: any semiotic or semantic theory must assume a distinction between sign tokens and ideal types which each repetition or "iteration" of a sign instances. But, Derrida argues, iteration cannot be repetition of stable semantic essence, for any significant particular can always be detached from its context and "grafted" into a new one in which it means something different. In Derrida's later work this undecidable logic assumes a broader ethical significance. Iterability implies that the text is *both* context-bound *and* transcends any *given* context, supposing "both that there are only contexts, that nothing exists outside context … but also that the limit of the frame or the border of the context always entails a clause of nonclosure. The outside penetrates and thus determines the inside" (Derrida 1988: 152). Any application of a moral or legal principle is thus potentially an act of reinterpretation or invention: "Each case is other, each decision is different and requires an absolutely unique interpretation, which no existing, coded rule can or ought to guarantee absolutely" (Derrida 2002: 251).

If iterability is a condition of thought or meaning as such, as Derrida argues, it implies that both have an open-textured temporal structure. The subject of thought, experience and intentionality is, accordingly, an "effect" of a mobile network of signifying states (or traces) structurally open to modification or recontextualization. Derrida's neologism *différance* captures this essential openness by capitalizing on the homonymy between the French verbs for differing and deferring. The identity or stability of the system of traces is differed-deferred because it is "vitiated by the mark of its relation to the future element" (Derrida 1984: 13–17).

The posthuman subject

For Hayles, the "autonomous liberal subject" she identifies with humanist theory is distinct from the conceptually ordered world in which it works out its plans for the good (Hayles 1999: 286). The *posthuman subject*, by contrast, is problematically individuated, because its agency is embedded and embodied in that world (*as per* active externalism) and because of the open, ungrounded materiality – or "iterability" – of language (Derrida 1988: 152; Hayles 1999: 264–5). The decentred posthuman subject is no longer sufficiently distinct from the world to order it autonomously as the subject of liberal theory is required to do.

But is this right?

Let's suppose, along with Hayles and other proponents of embodied cognitive science, that the skin-bag does not fix the boundary between agent and world or between the mental and non-mental. Nonetheless, even if thinking is a pattern of bodily and extra-bodily processes, this does not render thought or action less evaluable in terms of the rationality standards we apply to deliberative acts. As Badmington shows with respect to Descartes, rationality seems like a capacity that is manifested in our mental and *bodily functioning* (§1.5). An agent whose rational functioning depends on states of affairs beyond their skin is no less rational for all that. So even if the humanist subject is a swarm of bodily and extra-bodily agencies, this metaphysical dependence (or "supervenience") need not impair its capacity to subtend the powers of deliberation or reasoning liberal theory requires of it.[4]

If we accept his arguments, Derrida's account of general textuality nuances this picture by entailing limits on the scope of deliberation in the face of the "outside" or exception which infects any rule-governed system (Derrida 1988: 152).

But there is a difference between being ahead of oneself and being beheaded. The posthuman, in Hayles's critical sense of the term, *is not less human* for confronting the fragile and open-textured temporality of its cognitive and moral powers (see Chapter 6). The problem of how reason deals with the particular, the one-off, the exception, for example, is presaged in Aristotle's account of practical reason as well as Kant's account of aesthetic judgment: both insisting on the need for judgement to accommodate the singular or exceptional without resort to rules.

What Derridean deconstruction adds to this venerable tradition – as Martin Hägglund has emphasized recently – is the claim that these textual structures generalize beyond the sphere of the human. Concepts such as iterability and *différance* – (differing/deferring) originate in Derrida's readings of the philosophies of subjectivity but their sphere of application generalizes beyond this to all manner of machinic system: social institutions, living cells, computer programs and biological nervous systems

(Derrida 1978, 1998: 9; Hägglund 2008; Roden 2004b). But the fact that human subjectivity depends on structures shared by other biological or technical entities in no way levels functional differences between human and nonhuman (§1.4). After all, the traditional philosophical materialist will insist on a very similar thesis: "Humans are not special metaphysically. They are made out of the same fundamental particles, fields and forces that everything else is made out of." The fact that humans are made out of the same protons, quarks or electrons that everything else is made out of does entail that they are the same as everything else: to reason otherwise is to commit the fallacy of composition.

This is not to say that there is no merit in the model of the hybrid, open-textured self that Hayles and others present under the rubric of "the posthuman subject", or that it has no implications for "pre-critical" speculative posthumanism elaborated here. It does. I will argue in Chapter 4 that – far from being antithetic – critical and speculative posthumanism are complementary. A naturalistic position structurally similar to Derrida's deconstructive account of subjectivity can be applied to transcendental constraints on posthuman weirdness.

It will argue that a "naturalized deconstruction" of subjectivity widens the portals of posthuman *possibility* whereas it complicates but does not repudiate human actuality (Roden 2005, 2013). Understanding human agency in terms of iterability and *différance* leads to a moderately revisionary (but still interesting) account of what human rationality and agency consists in. But this leads us beyond the human by suggesting how rationality and agency depend on structures that are shared by nonhuman systems that may lack the capacities associated with human agency, or have other powers that humans do not enjoy (as an example of a loose application of the iterability argument to nonhuman systems, see my discussion of AI goals in §4.3).[5]

I conclude that the anti-humanist argument does not succeed in showing that humans lack the powers of rational agency required by ethical humanist doctrines such as cosmopolitanism. Rather, critical posthumanist accounts of subjectivity and embodiment imply a *cyborg-humanism* that attributes our cognitive and moral natures as much to our cultural environments (languages, technologies, social institutions) as to our biology. But cyborg humanism is compatible with the speculative posthumanist claim that our wide descendants might exhibit distinctively nonhuman moral powers (Roden 2010a; see Chapter 4, §§4.3, 8.2).

2.2 The technogenesis argument

Let's consider the second dismissal of SP – which begins from the claim that the human is "always already" technically constituted. In "Wrestling

with Transhumanism", Hayles argues that transhumanists are wedded to a *technogenetic anthropology* for which humans and technologies have existed and co-evolved in symbiotic partnership. Future transhuman enhancement would be technogenetic processes, according to this story; but so are comparable transformations in the deep past (Hayles 2011).

Human cultural and technical activity has, for example, equipped some with lactose tolerance or differential calculus without monstering the beneficiaries into posthumans (Laland *et al*. 2000)! Clark frames *the technogenesis argument* against posthumanism in his book *Natural Born Cyborgs* particularly well:

> The promised, or perhaps threatened, transition to a world of wired humans and semi-intelligent gadgets is just one more move in an ancient game … We are already masters at incorporating nonbiological stuff and structure deep into our physical and cognitive routines. To appreciate this is to cease to believe in any post-human future and to resist the temptation to define *ourselves* in brutal opposition to the very worlds in which so many of us now live, love, and work.
>
> (Clark 2003: 142)

Clark is famous for promulgating a variant of embodied cognitive science/active externalism known as the *extended mind thesis* (Chapter 5). Proponents of the extended mind thesis like Clark and Chalmers argue from a principle of "parity" between processes that go on in the head and any functionally equivalent process in the world beyond (Clark & Chalmers 1998).[6] The parity principle implies that mental processes need not occur only in biological nervous systems but in the environments and tools of embodied thinkers. If some chalk marks on a blackboard with which I record the steps of a length calculation make a cognitive contribution to my thinking, *they are as much part of my mental activity as the activation patterns in my brain.*

For Clark, humans are particularly adept at offloading processing demands onto external resources like written symbols and smart phones. Like Dennett (1991), Clark thinks these "hybrid mental representations" may account for our capacity to reflect on our own thoughts and thought processes – for example, via the use of embedding sentences such as "Joan believed that Bill is the culprit" (Clark 2008: 58; Bermudez 2002). If this is right, then skills like philosophical reflection and deliberation are the product of our technical-cultural activity rather than habits of bare brains. Rationality would already be due to a *cyborg coupling* of biological and cultural systems (though, as argued in the previous section, no less rational for all that – see Chapter 6, §6.5).

Clearly, if we restrict the evidence for the technogenesis argument to cases where technological change has not resulted in one species or group

splitting off from another, we are likely to infer that this is not liable to happen in the future. However, even allowing for this constraint, the fact that the game of self-augmentation is ancient does not imply that the rules cannot change.

Some pre-human divergence had to have happened in our evolutionary past and it is at least plausible – given Clark's cyborg anthropology – that technologies such as public symbol systems were a factor in the "hominization process" (see, in particular, Deacon 1997).

So there is evidence that one cultural-technological system may have been an evolutionary spur for the divergence of modern humans from their primate ancestors. Taking our cue from Churchland's proposal for centipedal cognition, it is conceivable that a cognitive augmentation of a similar order might accomplish a similar trick for posthumans were it to replace public language with a more flexible or potent medium of metarepresentation (§4.1). This is entirely compatible with Clark's hybrid account of biological/cultural representation since it involves the withering of a cultural component of one kind of hybrid mental state and its replacement with a new kind of hybrid mental state. Thus technogenetic anthropology is conceptually compatible with at least one scenario for the divergence of posthumans from humans. If technogenesis is conceptually compatible with one kind of posthuman/human divergence, it might be compatible with others (say, Vingean singularities).

Thus the technogenesis dismissal of SP invalidly infers that because technological changes have not monstered us into posthumans thus far, they will not do so in the future.

2.3 The materiality argument

Another of Hayles's objections to futurist visions of posthuman succession is their supposed denial or repression of the materiality of human embodiment and cognition: *the materiality argument* (this was discussed in §1.4). Computer simulations can help us understand the self-organizing capacities of the natural world, but this does not entail that any natural system can be fully *replicated* by a computational system that emulates its functional architecture or simulates its dynamics. The fact that cosmologists can simulate the evolution of galaxies with cellular automata does not mean that galaxies *are* cellular automata (Piccinini 2010: 279).

As we have seen, some transhumanist and speculative posthumanist scenarios presuppose a functionalist account of mind because they claim that minds could be fully emulated on computational soul engines (§1.3). This objection applies to a fairly restricted (if oft-cited) class of posthuman itineraries, however. SP is not committed to the singularity hypothesis; it is not committed to the possibility that humans could become digitized immortals.

It merely states that some of our wide descendants (human, machine, cyborg, etc.) might cease to be human in consequence of technological alteration (§1.4).

Perhaps, as Hayles hints in the opening of *How We Became Posthuman*, substrate neutrality collapses and beings differently embodied would also be differently minded (§1.4). If this was the case then uploading might not be possible, or if possible in some loose or extended sense, might not be consistent with the ethical humanity of the uploadee.

Thus the materiality of embodiment argument works in favour of the pre-critical posthumanist account (SP), *not against it*. It implies that weird morphologies can spawn weird mentalities.[7] On the other hand, Hayles may be wrong about embodiment and substrate neutrality. Mental properties of things may, for all we know, depend on their computational properties because every other property depends on them as well. To conclude: the materiality argument suggests ways in which posthumans might be very inhuman. It is, if anything, an argument for speculative posthumanism, not an argument against it (I will pursue this idea further in Chapter 3, §§3.1, 3.2).

2.4 The anti-essentialist argument

I turn, finally, to a dismissal that is perhaps implicit in some of the arguments considered above but which is worth considering for its speculative payoff. I refer to this as *the anti-essentialist argument*.

The anti-essentialist objection to SP starts from a particular interpretation of the disjointness of the human and the posthuman. This is that the only thing that could distinguish the set of posthumans and the set of humans is that *all posthumans would lack some essential property of humanness* by virtue of their augmentation history. An essential property of a kind is a property that no member of that kind can be without. If humans are necessarily rational, for example, then it is a necessary truth that if x is human, then x is rational.[8] It follows that if there is no human essence – no properties that humans possess in all possible worlds – there can be no posthuman divergence or transcendence.

This is a potentially serious objection to speculative posthumanism because there seem to be plausible grounds for rejecting essentialism in the sciences of complexity or self-organization that underwrite many posthumanist prognostications. Some philosophers of biology hold that the interpretation of biological taxa most consonant with Darwinian evolution is that they are not kinds (i.e. properties) but individual populations (see §6.4 for a fuller discussion). An individual or proto-individual can undergo a self-organizing process, but an abstract kind or universal cannot. Thus, the argument goes, evolution happens to species *qua*

individuals (or proto-individuals) not species *qua* kinds. To be biologically human on this view is *not* to exemplify some set of necessary and sufficient properties, but to be genealogically related to earlier members of the population of humans (Hull 1986).[9]

Clearly, if biological categories are not kinds and posthuman transcendence requires the technically mediated loss of properties essential to membership of some biological kind, posthuman transcendence envisaged by pre-critical posthumanism is metaphysically impossible.[10]

However, the anti-essentialist objection assumes that the only significant differences are differences in the essential properties demarcating natural kinds.

But why adhere to this philosophy of difference? The view that nature is articulated by differences in the instantiation of abstract universals sits poorly with the idea of an actively self-organizing nature underlying the leading edge cognitive and life sciences cited by Hayles, Haraway and other proponents of critical posthumanism. A view of difference consistent with self-organization would locate the engines of differentiation in those micro-components and structural properties whose cumulative activity generates the emergent regularities of complex systems (§5.4).

For example, we might adopt an immanent and particularist ontology of difference for which individuating boundaries are generated by local states of matter: such as differences in pressure, temperature, miscibility or chemical concentration. For immanent ontologies of difference, like the assemblage theory we explore in Chapters 5 and 6, the conceptual differences articulated in the natural language lexicons are asymmetrically dependent upon active individuating differences, not overbearing forms or transcendental subjects (Deleuze 1994; DeLanda 2002: 10).

In short: we can be anti-essentialists (if we insist) while being realists for whom the world is profoundly differentiated in a way that owes nothing to the transcendental causality of abstract universals, subjectivity or language.[11] But if anti-essentialism is consistent with the mind-independent reality of differences – including differences between forms of life – there is no reason to think that it is not compatible with the existence of a human–posthuman difference which subsists independently of our representations of them.

Looking forward

I have argued that critical posthumanists provide few convincing reasons for abandoning "pre-critical" posthumanism (SP).

The anti-essentialist argument just considered presupposes a model of difference that is ill-adapted to the sciences that critical posthumanists cite in favour of their naturalized deconstruction of the human subject.

The deconstruction of the humanist subject implied in the anti-humanist dismissal complicates rather than corrodes philosophical humanism – leaving open *the possibility of a radical differentiation of the human and the posthuman*. The technogenesis argument is just invalid. The materiality argument is based on metaphysical assumptions which, if true, would preclude only some scenarios for posthuman divergence while ramping up the weirdness factor for most others.

As we shall see in the following chapter, the main threat to SP is transcendental humanism since – in certain forms – it suggests that significantly powerful or self-optimizing forms of life would need psychologies or phenomenologies that conform to ours. Consequently, we shall explore transcendental humanism further in Chapter 3, considering some of its variants in the work of Kant (its originator), contemporary analytic philosophy, and phenomenology. In Chapter 4 I will develop some arguments that, I hope to show, unbind the constraints of transcendental anthropology.

However, this is just the beginning of an elaboration of SP and its philosophical implications. In Chapter 5 I will spell out the concept of what a posthuman is with greater precision in the form of the disconnection thesis. This will also develop an assemblage model of posthuman difference derived from DeLanda/Deleuze's account of immanent difference introduced in the last section. From there on we will be in a position to articulate the claim that posthumans would – for all their unbounded weirdness – be a kind of life.

Notes

1 The table specifies which operation the machine carries out when in a particular machine state (say, q_0) and a particular symbol is lying on the square currently being scanned. The table may, for example, specify that if the machine is in q_0 and a "0" is on the current square, then it should erase "0", replace it with a "1", move right, and enter another state (e.g. q_2). These simple "read", "erase", "write" operations can manipulate the contents of the tape, can generate an output corresponding to the value of a function when appropriately choreographed by the machine table – for example, the binary expression of a fraction.

2 Of course, even in classical cognitive science, the body and environment remain as "boundary conditions" for proper functioning of internal mental processes (Keijzer & Schouten 2007: 114).

3 This position has been explored in fiction in R. Scott Bakker's thriller *Neuropath* (2010) and in the work of philosophers such as Thomas Metzinger, whose views on what I call "dark phenomenology" will be considered in Chapter 4.

4 The notion of supervenience is used by non-reductive materialists to express the dependence of mental properties on physical properties without entailing their reducibility to the latter. Informally: M properties supervene on P properties if a thing's P properties determine its M properties. Suppose culinary properties supervene on physical properties: then if x is physically identical to y and x is tasty, y must be tasty.

5 We will also need to consider ethical complications arising from the temporal exposure of subjectivity to complex socio-technical systems in Chapters 7 and 8.

6 "Parity Principle. If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive

process, then that part of the world is (for that time) part of the cognitive process" (Clark & Chalmers 1998: 8). The PP is, in large part, a trivial consequence of functionalism. If mental states are individuated by their roles, then only the role and not the location of a state is relevant to it being mental. It is also open to the objection that the functional roles relevant to the individuations of cognitive processes are not borne by external representations (Rupert 2009). However, I will not adjudicate on this debate here since it has little impact on my thesis.

7 The argument may militate against the transhumanist dreams of virtual immortality alluded to above, but, as many have pointed out, this is a humanist or "hyper-humanist" scenario, not a posthumanist one (see Badmington 2003).

8 Another way of putting this is to say that in any possible world in which humans exist they are rational. Other properties of humans may be purely "accidental" – for example, 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.

9 David Hull points out that the genealogical boundaries between species can be considerably sharper than boundaries in "character space" (Hull 1986: 4). The fact that nectar-feeding hummingbird hawk moths and nectar-feeding hummingbirds look and behave in similar ways does not invalidate the claim that they have utterly distinct lines of evolutionary descent (LaPorte 2004: 44).

10 This objection is overdetermined because the possibility of successfully implementing radical transhumanist policies seems incompatible with a stable human nature. If there are few cognitive or body invariants that could not – in principle – be modified with the help of some hypothetical NBIC technology – then transhumanism arguably presupposes that there are no such essential properties for humanness.

11 For an excellent but somewhat neglected exploration of this idea in the context of contemporary anti-realism, see Farrell (1996).

3 The edge of the human

Introduction

In Chapters 1 and 2 I suggested that one of the distinctions between SP and transhumanism is that the former position allows that our "wide human descendants" could have minds that are very different from ours and thus be unamenable to broadly humanist values or politics. Vinge's speculations about the technological singularity provides an example of a posthuman weird tale (§§1.3, 1.4). But maybe there are constraints on posthuman weirdness that would restrict any posthuman–human divergence of mind and value. The possibility of an ontological catastrophe resulting from posthuman technologies will be reduced. SP would be correspondingly "bounded" because the scope for posthuman difference would be much less than some hope or fear.

But if there are significant constraints on posthuman possibility, how are we to find out what they are?

One potential source of information might be our current knowledge of the physical world or of information processing and computation. I will argue that, in the absence of actual posthumans, there is no evidence for significant constraints in these areas.

This will motivate the search for other "future-proof" constraints on posthuman possibility that we can know before evidence about the nature of actual posthumans is in.

The most coherent philosophical account of such *a priori* knowledge is to be found in the post-Kantian transcendental tradition. Some transcendentalists claim that worlds must be thinkable or experienceable for beings like ourselves; entailing that the form of the world is correlated with the structure of human subjectivity (Meillassoux 2010). If there are reality-constraints imposed by transcendental anthropology, these must also constrain the scope of posthuman weirdness. Transcendental humanism thus entails an *anthropologically bounded* SP. This chapter will attempt to formulate a plausible set of transcendental constraints of posthuman possibility with the help of central thinkers in the transcendental tradition such as Kant, Davidson, Husserl and Heidegger. This serves as a

prologue to Chapter 4 in which I will argue against anthropologically bounded posthumanism.

3.1 Bounds on posthuman possibility

I have argued that there are reasons for thinking that there could be nonhuman descendants of humans that have become nonhuman because of some technical alteration history. I have suggested that we should think of descent in "wide" terms in view of the likelihood that such relations will be technically mediated to an arbitrary degree. Finally, I have given reasons why the speculative posthumanist thesis SP does not commit us to a conception of the human that is unwarrantedly essentialist, which fails to reckon with the co-evolution of humans and technique, the nature of embodiment, or with the open-textured nature of subjectivity (§2.4 – this position will be honed considerably in Chapter 5). Some of the objections to SP promulgated by critical posthumanists have been shown to rest on confusion between SP and transhumanism, or on precipitate claims about the implications of theories such as deconstruction or embodied cognitive science (§2.1).

However, even given SP – given *that there could be posthumans* – it does not follow that every conceivable posthuman is a possible one. If some posthumans are impossible, that is presumably because there are real constraints on the kinds of beings that are possible in this world. If we could discover some of these constraints, we could begin to narrow down the field of our possible technological successors. If any of these are necessary constraints, we will also be able to say something about what posthumans would have to be like. Thus exploring potential constraints on posthuman life seems like a method for developing a more positive account of the posthuman even in the face of their dated nonexistence.

We can make this idea more precise by considering the collection of physically and technically possible histories whereby posthuman wide descendants of humans could emerge on this planet – *posthuman possibility space* (PPS).

Recall that wide descent is technically mediated to an arbitrary degree. So PPS could include many different paths to posthumanity corresponding, perhaps, to prospective NBIC technologies. The only thing that these itineraries need share is that they are the result of feasible technologies. For example, if faster than light (FTL) travel is impossible in our universe – as general relativity suggests – then no posthumans will be able to FTL. If machine AGI is – for whatever reason – impossible, then PPS will not include any paths to posthumanity involving AGI, and so on (§1.2).

As yet we know very little about PPS. We saw in Chapter 1 that it is legitimate to conceive becoming not human in essentialist or anti-essentialist

terms. That is, we may think that there are properties necessary to being human that posthumans lack. But we may also deny essentialism and consider human–posthuman differences to be historically emergent relations of some kind.

Though we know nothing about them, as yet, we can think of these possible posthumans as corresponding to the posthuman states of the world in PPS.

For all we know, PPS contains nothing at all.[1] This would mean that posthumans – however conceivable – could not occur in a world like our own. Alternatively, it may be thronging with inhuman histories. Perhaps you will occupy one of these histories, either because you will become posthuman or you will encounter them in your future.

At this point in our investigation, it need not concern us which of these alternative scenarios is true. All that we need assume is that PPS exists and is either empty or nonempty.

3.2 Natural constraints on PPS

Whether PPS is empty or nonempty depends on what is possible in our world or any world whose structure or laws are similar. I will refer to this broad notion of possibility as "natural possibility".

However, SP is primarily concerned with technological possibility. A technologically impossible posthuman would be impossible *period*. Not every naturally possible state need be technologically possible. For example, sustained nuclear fusion might only be producible by gravitational confinement in a star.

It might seem that some technological possibilities can be discerned *a priori* – by consulting reliable conceptual "intuitions" about the extendible powers of current technologies. For example, a being like Skynet – the genocidal military computer in James Cameron's *Terminator* films – seems a plausible occupant of a PPS timeline; whereas Sauron, the supernatural dark lord of Tolkien's *Lord of the Rings*, does not. However, since the work of Saul Kripke in the 1970s many philosophers have come to accept that there are *a posteriori* natural possibilities and necessities that are only discoverable empirically. That light has a maximum velocity from any reference frame upsets common-sense intuitions about relative motion and could not have been discovered by reflecting on pre-relativistic concepts of light (Fine 2002).

Claims about hypothetical technological possibility may be as vulnerable to refutation as naïve physics. States like the US and China employ computers to co-ordinate military activities so a Skynet seems the more plausible posthuman antagonist. But the fact that there are computers but no dark lords does not entail that their capacities could be extended

in any way we imagine. Light bulbs exist as well as computers, but maybe a Skynet is no more technologically possible than Byron the intelligent light bulb in Thomas Pynchon's fabulist novel *Gravity's Rainbow* (see §5.6).

This prescription for epistemic modesty is supported by a recent study of past predictions about the future of artificial intelligence. Stuart Armstrong and Kaj Sotala have collated expert predictions regarding the nature and timeline of key advances in AI, suggesting that these are contradictory and barely distinguishable from the predictionsArmstrong & Sotala 2012). On this basis, they argue that it remains unclear whether revolutions in AI development will require just more hard work and money ("grind") or some new "insight" (pp. 64–5).If it is the latter, any predictions about shape and nature of the machine minds of the posthuman future are liable to be highly error-prone.

Perhaps drawing on *a posteriori* physical possibilities that have been discovered as our guide to technological possibility will give us clues about some possible occupants of PPS as well as some possible constraints on membership. Anders Sandberg (1999) suggests that any intelligent system will need to store and transform information in a physical medium.[2] Perhaps the physical limits of data processing will apply to all denizens of PPS.

There are physical constraints on the data that can be stored in a given kind of medium, and constraints on the speed and accuracy with which that information can be transformed. According to Eric Drexler, computer memories that code bits at the atomic level may enable data to be stored at a density of approximately a billion million million bits per cubic metre if the storage medium approached the density of diamond (Drexler 1992, cited in Sandberg 1999). MP3 players and smartphones are technically possible because hard disks can store in the order of ten billion bits per cubic metre (Walter 2005). Thus current data storage operates many orders of magnitude below the theoretical limit available in ordinary matter.[3]

Any physical limits on information storage density and processor speed will presumably constrain all occupants of PPS unless fundamental physical laws or powers can change. These numbers suggest that current information processing capacities may be many, many orders of magnitude below the maximum allowed in nature.

Even apart from physical constraints of the kind just mentioned, one can ask what kinds of system can perform such computation. Jiří Wiedermann cites results from his research suggesting "amorphous systems" made up of independent units forming into ad hoc networks (e.g. tiny nanomachines or genetically engineered microbes) could have universal

computing power: that is, in principle, such a system could compute any program that a universal Turing machine could compute (Wiedermann 2012: 83–4; §2.1).

However, as noted above, not all programs that are computable (in the Turing sense) are *tractably computable*. Some computations – like generating all the sequences of sixty different things – may be simple enough to program, but still take longer than the lifetime of physical universe to complete (Biermann 1997: 374–5). Inferences that take account of holistic properties of representations may – for all we know – be computable in a purely mathematical sense[4] but the relevant programs may not be performable under real world time constraints (§2.1; Horgan & Tienson 1994; Eliasmith 2002).

Given these uncertainties about the computational basis of mind, it is hard to derive strong *a priori* constraints on the contents of PPS from constraints on physical information processing or efficient computation. The results reviewed here suggest only that there may be natural scope for information processors that are much faster and fatter than humans.

Are there any philosophical principles that suggest how the space of possible minds might be constrained by factors over and above the physical constraints considered?

We have already considered two opposing accounts of the relationship between embodiment and mind in our discussion of Hayles's materiality claim in the last chapter. She asserts that a computational emulation of a being with a mind – which matches the causal-functional roles of all that being's components (e.g. neurons, inter-neuronal connections, and chemical modulators) – could never duplicate its mental states. The opposing view, of course, is that mental states could be duplicated if the being's components were emulated at a sufficiently fine grain.

Hayles fails to support the materiality claim; but *it is* supportable. The argument assumes a metaphysical distinction between the *dispositions* or *powers* of a thing and the way in which those dispositions are manifested in disparate contexts. As John Heil points out, we ordinarily describe a disposition by its manifestation – for example, elasticity, sonorousness, solubility, irritability, though the disposition itself could have manifested differently in a different environment.[5] How a power gets to manifest itself thus depends on triggering conditions (see Heil 2003: 83). Thus it does not follow from the fact that we describe dispositions in terms of what they do that they are just the sum of what they do; or even that powers claims are equivalent to conditional claims about possible manifestations (Molnar 2006: 83–92).

It follows that a system whose input–output behaviour was relevantly similar to a human might still have different powers to a human by virtue of its different composition and structure.[6] For example, suppose

that my brain and body were emulated to a fine level of detail on "computronium" – a form of programmable matter like the materials anticipated by Drexler – which encodes information at enormous densities and allows information processing with speeds several orders of magnitude greater than any current information processor on Earth. Call this functional duplication $David_C$.

Let us assume a substrate neutrality sufficient for consciousness, intentionality and other key mental properties to be realized on non-organic as well as organic substrates (§1.3). $David_C$ might start out with mental states with a similar structure and content – the same beliefs, desires, etc. – as my biological duplicate. However, $David_C$'s computronium body is capable of thinking a billion times faster than those of his biological friends. So, rather than engage in tiresome social interaction with them, he delegates his human public relations to a further (human-speed) emulation $David_E$ running on a tiny volume of his computronium core. Even if $David_C$ starts out as a functional duplicate of David Roden, his powers would be significantly different by virtue of his accelerated thought processes. Once $David_C$ includes $David_E$ as a public relations module, we no longer have a system that is plausibly identical to David Roden at a functional or computational level. Because of his $David_E$ homunculus he could outwardly behave much like I behave in human social contexts (responding the same way to the same questions or prompts from his human friends) but his appearance would be deceptive.

If the time-window required for thinking is relevant to making assessments of psychological similarity, then people running on radically different computational substrates are unlikely to be psychologically similar because their powers – including their ability to alter their own functional structure – will be different (Chapter 6). We can only produce psychologically similar copies of ourselves on alternative substrates where the powers of those copies are relevantly similar (call this the *power-identity assumption* – PIA).

The PIA implies that creating a functional duplicate of a human on a more or less efficient computational substrate would not necessarily duplicate their psychology (see Eliasmith 2002). If this is right, people-emulations on digital computers or other non-biological substrates may be different in ways that we cannot yet predict because the powers that they have in virtue of differences in computational speed or efficiency are not ours.

As observed in the previous chapter, the metaphysical constraint represented by PIA does not make the future shapes of mind any easier to predict – if true, it is metaphysically and ethically salutary but provides no hard information about PPS. It suggests that predicting posthuman lives is harder because minds embodied in different substrates will not be psychologically identical (see §5.6).

To be sure, if fine differences in substrates were always relevant to gross behaviour, computer modelling would be impossible. Reality would not be decomposable into chunks whose stability in the face of lower level differences can be assumed (DeLanda 2011: 13–14).[7] However, short of emulating a sophisticated mind on some actual computational substrate, it remains unclear whether powers-differences between biological humans and their uploaded or emulated counterparts would be negligible in technologically feasible cases.

3.3 Kant and transcendental humanism

If we seek bounds on posthuman weirdness, perhaps we should look for knowledge that is truly *a priori* knowledge: non-trivial information about the future that can never be disconfirmed by subsequent evidence. By "non-trivial" I mean that such knowledge would have to ascribe properties to posthumans not already implied in some concept of posthumanity we have constructed. If we specify in advance that posthumans are nonhumans, it follows trivially that no posthuman (thus conceived) will be human.

The claim that we could have non-trivial *a priori* knowledge of the nonhuman world runs contrary to the minimal realist assumptions that: 1) the nonhuman world exists and 2) its nature or structure is independent of our representations of it (Devitt 1991: 32–3). If the world is mind-independent, as the realist claims, it can always be other than minds represent it as being.

Descartes' evil demon thought experiment is, of course, one way in which we can make the realist idea of the autonomy of reality vivid (§1.4).

Kant – who believed that the *a priori* is necessary for science – bites this bullet by rejecting the realist construal of knowledge as bearing reference to a mind-independent reality, as well as the sceptical problematic that arises from it. His startlingly original conclusion is that *objects must conform to their representations in certain ways to be knowable at all*. Meanwhile subjectivity – in particular, our experience of being a unitary self over time – is only possible if it refers to objects (Kant 1978: B133). Mind and world are thus not mutually independent, as Descartes thought, but inseparably related. "Transcendental" knowledge is the part of philosophy that is concerned with the subjective conditions for the human–world relation (B25).

Kant holds that knowledge would not be possible unless the transcendental subject already interprets its experience as about a common, objective world. Descartes thought that knowledge could be founded by retreating to an inner world whose ideas are clear and immediately evident to the experiencer (§1.4). But Kant denies that we could be aware of mental events prior to giving some of them objective content. We cannot

be aware of changes in our "inner" states, according to Kant, unless we can identify something outside ourselves to which these occur (1978: A181/B224–5; Guyer 2006: 106–08). This objective purport comes about by *synthesizing* our disparate sensations into conceptually ordered experiences using high-level concepts known as "categories".

For example, to observe that a melting block of ice persists across the changes in its sensory appearance, I apply the concept *substance*. To judge that these changes result from exposure to a heat source, I use the concept *cause.* The first synthesis makes my experience of change possible by attributing it to a persistent thing. The second gives succession a determinate order in time, since if a relation between events is causal, one follows as a necessary consequence of the other (Longuenesse 2005: 25, 158). Without this categorical ordering, these experience could only concern disparate sensations or features; which is just to say that they would not be experiences of a being aware of its mental life or the world around it (see §4.1).

This provides the transcendental warrant for the Kantian argument. If the application of the categories is a condition of self-awareness in human subjects, then from the fact that we are self-conscious we can infer that the empirical world of ice blocks, galaxies and prospective posthumans does indeed conform to human categories.

Kant accepts that there are mind-independent things-in-themselves – *noumena* – that do not conform to this human-imposed structure. However, he denies that we have knowledge of them because we only have access to the empirical world generated by our categorical activity. These are ineluctably ordered in a single space and time, which, he argues, are the outer and inner forms of human sensory experience rather than mind-independent "things-in-themselves" (Kant 1978: B37, A23). So the properties of things that we can know about are not properties of *noumena* but properties of *phenomena* – categorically ordered things in space and time.

For transcendental synthesis to occur there must be an entirely non-representational relationship between categories and the content of experience, since the sensations organized by synthesis have no objective content of their own – the category cannot match or conform to the sensory manifold it unifies. Kant's explanation is that the mind applies a "schema", a rule for ordering experience that conforms to a concept (Kant 1978: A140/B180). Truth, then, is not a correspondence between an inner mental realm and an outer nonmental one, as it is for the realist. For the judgement "The popcorn has been cooked" to be true there must be both empirical cooking and transcendental cooking. The popcorn needs to have been heated. But it must also be possible for any human observer to generate an experience that would allow us to concur with this claim (Braver 2007: 50).

As Braver points out, this interpretation of truth as intersubjective agreement would fail to justify belief in a unique empirical reality unless transcendental subjectivity is stable across observers. Suppose you have an alien cognitive nature that employs different categories to connect your perceptions. None of the judgements I could make about my world could be true of your world – or vice versa. We would experience distinct worlds without common points of reference. Attributing a common human nature to subjects means humans cook up a shared phenomenal world even if they do not hook up to noumena (*ibid.*: 49).

Thus, as advertised in Chapter 1, Kant's position is a drastically new form of humanism: transcendental humanism. It implies that humans are distinctive in virtue of being the transcendental architects of the human–world correlation.

3.4 Pragmatism and phenomenology

Philosophers who followed in Kant's transcendental footsteps – for example, Hegel, Husserl, Heidegger and Davidson – have advanced recognizably transcendental claims which identify the conditions of possible knowledge or meaning with facts about human subjectivity. However, most of these successors have argued that the idea of the noumenal beyond is unintelligible. The only conception of reality we have is of one we can know and access. For example, modern phenomenologists claim that transcendental philosophy must devote itself to the investigation and understanding of "what appears" – bracketing speculative assumptions of the kind found in realist metaphysics. Phenomenology does not treat appearances as mental copies of mind-independent noumena (Heidegger 1962: 51; Braver 2007: 182–5). For Heidegger and for Husserl the phenomenon (that which appears) "shows itself in itself".

Now, making sense of mind-dependence turns out to be as hard as making sense of mind-independence. Kant's original account of transcendental subjectivity is notoriously problematic. For example, Kant admits that he cannot explain how a schema imposes form on an intrinsically formless manifold of sensation.

However, modern pragmatism, like phenomenology, seems well placed to avoid this "scheme content" dualism (Davidson 1984: 183–98). Pragmatists are committed to the claim that conceptual and intellectual powers are grounded in our practical abilities rather than in relations between mental entities and what they represent (Brandom 2006). So while pragmatists buy into the Kantian claim that concepts are cooks, not hooks, they are leery of all the "transcendental psychology" that goes with it (but see §4.1). Modern pragmatists argue that conceptual understanding is exhibited in our practical grasp of public norms of reasoning

rather than in mysterious agencies of the mind (Brandom 2001: 6; Sellars 1963).

This results – arguably – in a significant gain in intelligibility and practical import because concepts are now implicated in human social life (Rorty 1980; Levine 2010: 582–3). Language is not a medium for expressing our inner selves or outer realities, but a social matrix that can be revised by the proposing of new paths through the space of reasons (Rorty 1989: 18–19).

3.5 Discursive agency

Pragmatism is an attractive and widely held doctrine because it promises to elucidate difficult notions like meaning and truth in terms of human activities rather than problematic transcendent metaphysics. It is precisely because pragmatism elaborates transcendental humanism plausibly that we need to consider its implications for posthuman possibility.

The first consequence of the pragmatist idea that language is the matrix in which we cooperatively form and revise reasons is what I will call the *discursive agency thesis* (DAT). DAT asserts that agents must have the capacity to use public language in social contexts. If true, DAT would be a significant *a priori* constraint on SP since, until now, we have not assumed that posthumans would have to be language-using or social.

The argument for discursive agency falls out of the broader pragmatist claim that the rational coordination of beliefs, desires and intentions is a social skill:

1) An agent is a being that acts for reasons.
2) To act for reasons an agent must have desires or intentions to act.
3) An agent cannot have desires or intentions without beliefs.
4) The ability to have beliefs *requires a grasp of what belief is* since to believe is also to understand "the possibility of being mistaken" (metacognitive claim).
5) A grasp of the possibility of being mistaken is only possible for language users (linguistic constitutivity).

So a being that lacks the capacity for language cannot be an agent.

Psychological states such as beliefs, desires and intentions (along with hopes, wishes, suppositions, etc.) are commonly referred to as "propositional attitudes" because they are expressed as an attitude towards the content of a declarative sentence embedded in a "that" clause (for example, I believe that *Lima is the capital of Peru*. You hope that *Manchester United will win the European Champions League*).

Many philosophers take propositional attitude psychology to be the core conceptual framework with which humans predict and interpret each

other in interpersonal life. Davidson accepts this, but argues that the social activity of attributing and evaluating beliefs serves a function comparable to the transcendental categories in Kant (§3.3). It allows us to understand objectivity. For Davidson, our grasp of objectivity and truth falls out of the "triangular" situation in which "two (or more) creatures each correlate their own reactions to external phenomena" (Davidson 2001b: 129).

This idea informs the first of the two strong assumptions in the DAT argument, which I have referred to as "the metacognitive claim" because it asserts that having mental states *about* mental states (in this case, a concept of what a belief is) is essential to having beliefs. For Davidson, belief is an attitude of "holding" true some proposition: for example, that there is a cat behind that wall. But if belief is *holding true* it seems to require a grasp of truth and falsity and thus of belief itself. Thus we cannot believe anything without grasping that others could have beliefs about the same topic (Davidson 1984: 170; 2001b: 104 – though see §4.1).

This leads us to Premise 5 (linguistic constitutivity). For Davidson, the intersubjective triangle between the subject, another person and the world is only accessible to a being which can actively attribute beliefs (Davidson 2001b: 129; Briscoe 2007: 140–41). But for triangulation to occur, there must be a common framework in which two or more beings can compare and evaluate their respective takes on the world through dialogue. According to Davidson, language provides this dialogic framework for representing differences, similarities and relations between beliefs:

> Our manner of attributing attitudes ensures that all the expressive power of language can be used to make such distinctions. One can believe that Scott is not the author of *Waverley* while not doubting that Scott is Scott; one can want to be the discoverer of a creature with a heart without wanting to be the discoverer of a creature with a kidney. One can intend to bite into the apple in the hand without intending to bite into the only apple with a worm in it; and so forth. The intensionality we make so much of in the attribution of thoughts is very hard to make much of when speech is not present. The dog, we say, knows that its master is home. But does it know that Mr Smith (who is his master), or that the president of the bank (who is that same master), is home? We have no real idea how to settle, or make sense of, these questions.
>
> (Davidson 1984: 163)

However, if language is necessary for belief, and beliefs are necessary for agency, Davidson's position is both transcendentally humanist and strongly anthropocentric. Language is a condition of possibility for

rationality, objective thought and agency (rationality is a "social trait", Davidson claims).

Non-language using animals may be sentient but, as Robert Brandom puts it, they are not *sapient* (Brandom 2001: 157; Wennemann 2013: 47). They cannot be answerable to reasons, identify themselves as the thinkers of several thoughts over time, or have any grasp of being in the world of objective things.

Davidson's account of linguistically mediated rationality thus reformulates the transcendental humanist thesis that humans constitute the world and are not just things in it. This process is now seen in terms of intersubjective dialogue rather than Kant's hoary transcendental psychology. We have a relationship to a world only if we trade propositions in common linguistic coin. Davidson does not think that this communicative structure requires common linguistic conventions; but it does require the capacity to impute reasons to others' verbal behaviour, even where, as with Lewis Carroll, James Joyce or Mrs Malaprop, their speech deviates from the norm (Davidson 1986). Humans are animals with the social capacity to be "gripped" by concepts and norms of reasoning (Brassier 2011).

The DAT implies that the occupants of posthuman possibility space (PPS) will need to be subjects of discourse if they are to be agents. Daryl Wennemann makes this assumption in this book *Posthuman Personhood*. Wennemann adopts the traditional Kantian idea that agency consists in the capacity to justify one's actions according to reasons and shared norms. For Wennemann, a person is a being able to "reflect on himself and his world from the perspective of a being sharing in a certain community" (Punzo 1969, cited in Wennemann 2013: 47). And this is a condition of posthuman agency as much as of human agency:

> In a posthuman age, the moral community is constituted by all beings of a kind that are capable of moral reflection and thus agency. Human beings are one such kind. There may be other kinds as well (computers, robots, aliens). So, to identify morally with the members of the moral community in a posthuman environment is to identify morally with all beings of a kind capable of agency.
>
> (Wennemann 2013: 49 – citation modified[8])

It follows that posthuman agents will need concepts, desires and intentions expressible in the social idiom of sentences: the full panoply of propositional attitudes. This is a problem if we still entertain the idea of the radically *weird* posthumans that Vinge holds out for in "The Coming Technological Singularity". For the DAT implies that we can know *a priori* that the *structure* of posthuman thought and agency would be

discursive, even if posthumans have strange bodies or social habits. Such differences seem superficial in the light of the deep transcendental structures entailed by the DAT. After all, humans differ in gender, skin colour and physiognomy; and people from different cultures often live according to contrary conceptions of the good life. These differences are politically important in isolated cases where conceptions of the good life clash, or where sexists, racists and xenophobes make them so; but they do not run particularly deep. If so, it is hard to see how mere differences in appearance of substrate could have distinctive metaphysical or moral import attributed to SP, let alone carry the eldritch promise of the radical weird.

3.6 Pragmatism and anti-realism

Unlike Kantian transcendental philosophy, pragmatism seems compatible with the mind-independent *existence* of the world and thus with one half of realist claims that the nonmental world is existentially independent of our minds and has a nature independent of our cognitive activity (§3.3; Devitt 1991). After all, it is committed to practices and thus to their material supports like pencils and power lines. However, we shall see that pragmatism is not compatible with an independent nature because it implies that the way the world is articulated or "contoured" depends on discursive practices.

If so, pragmatism implies a second *a priori* constraint on PPS. If posthumans are agents and subjects of discourse (from DAT) their world is discursively articulated, even if the practices by which they carve up the world might differ from ours. In the next section I use considerations of mutual interpretability to argue that a pragmatist world is not an aggregate of things (lumps) but an open-textured "horizon" differentiated by the activity of speaking (human, posthuman, alien) subjects. This is clearly a form of what Meillassoux entitles *correlationism* (§1.4). It implies that posthuman and human lives would be co-correlated with the *same* world horizon.

In modern analytic philosophy the rejection of the second plank of realism inaugurated by Kant's turn is referred to as "anti-realism". We lack space to consider all the varieties of pragmatic anti-realism, so in what follows I'll consider a relatively clear cut anti-realist doctrine that Hilary Putnam refers to as "internal realism" (IR).

First, to understand why IR is not really realism, we need to introduce Putnam's influential analysis of traditional realism – or *metaphysical realism* (MR).

MR is not one thing, according to Putnam, but a bag of interrelated claims about the mind–world relationship. The key components of MR

are 1) the *independence thesis*; 2) the *correspondence thesis*; 3) the *uniqueness thesis*.

The *independence thesis* states that there is a "fixed totality of mind-independent objects" (the world) (Putnam 1981: 49).

The *correspondence thesis* states that there are determinate reference relations between bits of language or mental representations and the bits of the world to which they refer.

The *uniqueness thesis* states that there is a single theory whose sentences correctly describe the states of these objects. This implies a singular correspondence between the terms belonging to this theory and the objects and properties that they refer to.

MR is a package deal. The correspondence thesis needs objects (independence) for the mental and linguistic bits to correspond to. There must be mind-independent objects for there to be a single correct way in which the One True Theory corresponds to the world (uniqueness entails independence and correspondence).

Putnam presents this idea in terms of a branch of mathematical logic known as "model theory". Model theory is an abstract way of understanding the links between formal languages and the "world" their sentences are about. A model is a set of objects. These can be material things like cats and elementary particles, or abstract objects like sets or numbers. In model theory, a formal language (a collection of uninterpreted symbols organized by a grammar and rules of inference) is given an interpretation by assigning elements or subsets of the model to its symbols. This assignment is called an *interpretation function*.

In model-theoretic terms, MR is just the claim that there is a unique description of the world *hooked up to that world by a single interpretation function*. The singularity of the interpretation function is crucial because if there was more than one way of interpreting the terms of the One True Theory there would not be a single correct description of the world. Uniqueness (thus MR) would fail.

What virtues could help us distinguish the One True Theory from competitors? According to Putnam, it would need to satisfy the "operational constraints" that ideally rational inquirers would impose on such a theory. If one imagines science progressing to an ideal limit at which no improvements can be made in its explanatory power, coherence, elegance or simplicity, etc., then the One True Theory would have to be as acceptable to ideally rational enquirers as that theory (Putnam 1981: 30, 33).

However, Putnam argues that even if we were to find a theory that satisfied these constraints at the ideal limit, it would be possible to find another with the same scientific virtues just as true as it is and equally as consistent with our practices of speech and justification. Thus – failing

other facts that could distinguish an ideal theory as God's Own – uniqueness fails and, with it, MR.

Putnam's argument against MR is supported by a theorem of model theory. This states that for any language whose referring expressions are mapped onto objects in a range of possible worlds (each associated with a "fixed totality of objects") by an interpretation function I, there will always be a second interpretation function J that maps the same expressions onto different objects picked from the same world which preserves the truth/falsity of all the sentences of the language (pp. 217–18). Thus the general term "cat" might refer to the set of cats under I but to the set of cherries under J. This will not affect the truth value of a sentence like "Fred is a cat" so long as the new interpretation J maps "Fred" onto a member of the set of cherries.[9] So "Fred is a cat" will remain true under J but will mean something different.

Thus even a theory that satisfies the ideal of operational virtue is convertible to a second *equally good* theory in the same world by shuffling around the meanings of its symbols. The second theory is equally good because the truth values[10] of all of its sentences in the language are retained under J. Any observation sentence that supports the theory under I will support it under J. Any true prediction that follows from the theory under I will follow under J, etc.

If this is right, for every true theory of the world, there is at least one other true theory. Metaphysical realism fails because uniqueness fails: even God, it seems, cannot have a preferred way of describing the world.

Some might object that uniquely intended interpretations can be mentally imposed by our beliefs or ideas. So on this account theories are not interpretable formal languages but intrinsically meaningful thoughts in our minds. This might appear to block the conversion of the first good theory to another because the model-theoretic argument applied to languages, not to thoughts.

However, it can be objected that this response just assumes there can be a "magic language" of self-interpreting mental signs (Wheeler 2000: 3). But the same argument can be applied to a thought as can be applied to the terms of a formal language so long as we assume that thoughts are expressed in some medium or other (e.g. "mentalese"). Thoughts, on this account, are just mental signs. Mental signs do not interpret themselves and wear their meaning on their sleeves any more than anything else (Putnam 1983: 207). The problem is not that there are no correspondences between signs and things. There are too many to secure a unique interpretation-independent meaning for even the best of theories (Putnam 1981: 73; Moran 2000: 78). Thus appealing to "inner" or mental signs to fix the intended meanings of our theories reruns the problem of indeterminacy all over again.

So, for Putnam, the possibility of gerrymandering new interpretations from old implies that MR is false. There is no single theory that uniquely describes the realist's mind-independent world and thus no non-interpretation-relative way the world is.

Rather than aspiring to the idealized God's Eye View of metaphysical realism, Putnam argues we should recognize that truth, reference and objectivity are properties that our languages have because of "our" social practices of inference, confirmation and observation (pragmatism again). To assert "'Cow' refers to cows" is not to make a claim about some determinate relationship between word and world but to make a normative claim about how a competent speaker of English should use "cow" around these parts (Putnam 1978: 128, 136). This does not reflect some metaphysical insight into the mind–world relationship but our tacit grasp of a particular form of human life (p. 137).

Before going on to consider the implications of this anti-realist position for pragmatist conceptions of understanding and worldhood, it should be pointed out that there are lines of attack open to those who wish to retain some version of realism. Some, like Fodor, argue that there are semantic truths about what refers to what that are interpretation-independent (Fodor 1990; Hale & Wright 1999). Thus the existence of alternative interpretations of a theory does not entail that they are equally acceptable. Perhaps the world is differentiated in one way rather than another and a One True Theory would hook onto those differences regardless of whether alternative referents were assignable to its sub-sentential terms. For example, any theory could be given a model consisting of sets and set-theoretical structures. It does not follow that set theory is just as good an interpretation of some part of physical theory as one which posits elementary particles or fields.

More interestingly, one might claim that MR is just a logician's caricature of realism to which the metaphysical realist need not be committed. For example, Devitt denies that realism is really committed to uniqueness – the view that there is exactly "one true and complete description of the world" (Devitt 1984: 229). We might also demur from the assumption that the world consists of objects or only objects that enter into semantic relationships with bits of language or mind. Structural realists, for example, argue that reality *is* structure and that this is precisely what approximately similar theories capture – regardless of their official ontological divergences (Ladyman & Ross 2007: 94–5). Some speculative ontologies deny the correspondence assumption, holding that the world contains entities that cannot be fully represented in any theory: for example, powers, Deleuzean intensities, or Harman-style objects (see §6.6).

Perhaps, the correspondence assumption just replicates the Kantian view that entities must conform to linguistic modes of representation

(§3.3). If, as I argue in Chapter 4, we have no secure grounds to make such wide-ranging transcendental claims, then the scope of Putnam's argument can also be deconstructed from within.

It is not clear, however, that any of these responses are open to pragmatists, for whom formal semantic facts like those captured in model theory are ultimately mystified expressions of our communal or idiomatic practices. So I want to consider how pragmatists might square the failure of uniqueness with the requirement that communication and interpretation take place in a shared world. Is the pragmatist entitled to a unique world and, if so, what manner of world is it?

3.7 From pragmatism to phenomenology: Davidson, Husserl and Heidegger

Putnam's anti-realism implies that we have no better conception of what the truth of a sentence or belief consists in other than its acceptance by ideally rational inquirers when all the evidence is in. Davidson is on record as rejecting this epistemic conception because it appears to relativize truth to languages (Davidson 2001b: 186–7). However, whether this charge is justified is irrelevant for our purposes. Davidson, as we saw in §3.5, is committed to the claim that each believer has the concept of an objective world about which she or others can have "true" or "false" beliefs:

> Communication depends on each communicator having, and correctly thinking that the other has, the concept of a shared world, an intersubjective world. But the concept of an intersubjective world is the concept of an objective world, a world about which each communicator can have beliefs.
>
> (Davidson 2001b: 105)

Putnam's conception of truth at the limit of enquiry may (or may not) be weaker than Davidson's, but *it has to furnish a conception of an intersubjective world*. For grasping it must allow me to understand that my beliefs may not be confirmed under ideal conditions by smarter creatures with access to more evidence. The concepts of truth and intersubjectivity are thus as interconnected in Putnam as in Davidson.

This is also the case with Davidson's account of meaning. Davidson thinks that the best way to illuminate the notion of meaning in philosophy is to characterize what we know that allows us to interpret fellow language users. Davidson claims that an interpretation can be represented using the logical machinery devised by Alfred Tarksi to derive the truth conditions for any sentence in a formal language (Davidson 1984: 17–36).

The argument for the adequacy of truth theories for encoding our linguistic competence is justified in terms of a thought experiment which envisages a field linguist interpreting an alien language from scratch: the ideal of *radical interpretation* (pp. 125–37).

Davidson argues that the criterion of hermeneutic success in radical interpretation is that the truth theory correctly predicts circumstances of utterance for arbitrary utterances by stating what would make them come true. For example, a correct truth theory for English would state (in its language) that the sentence "Snow is white" is true if and only if snow is white.

If Davidsonian truth theories capture a competent speaker's grasp of meaning, no part of a language – for example, the predicate " … is white" – can be understood unless we understand the truth conditions of all the sentence in which it occurs ("Snow is white", "Cotton is white", etc.). But the connections ramify, since these sentences will also depend on other parts ("Snow") whose meaning requires still other truth conditions to be spelled out ("Snow is a form of water").

The upshot of this is the holist thesis (familiar from Saussure's structuralism and some versions of inferential role semantics) that the meaning of any term in a language depends on its interrelationships with all the other terms in a language (Davidson 1984: 21; Evnine 1991: 120; Brandom 2007).

Davidson's meaning holism also implies *psychological* holism for, as the DAT implies, sentence meaning and psychological content are interdependent (Malpas 1992: 86–7). Our capacity for belief and agency depends on our capacity to interpret utterances in the light of novel speech behaviour. No belief content and no content of any other psychological state can be fixed in isolation:

> If someone is glad that, or notices that, or remembers that, or knows that, the gun is loaded, then he must believe that the gun is loaded. Even to wonder whether the gun is loaded, or to speculate on the possibility that the gun is loaded, requires the belief, for example, that a gun is a weapon, that it is a more or less enduring physical object, and so on. There are good reasons for not insisting on any particular list of beliefs that are needed if a creature is to wonder whether a gun is loaded. Nevertheless, it is necessary that there be endless interlocked beliefs. The system of such beliefs identifies a thought by locating it in a logical and epistemic space.
>
> (Davidson 1984: 156–7)

One of the important (and contested) conclusions that Davidson derives from the metatheory of radical interpretation is that a prospective interpreter must adopt a regulative principle of charity by assuming that speakers of the language under interpretation are rational and not systematically mistaken. Were speakers systematically duped about their world, their

public utterances would reveal nothing about their truth conditions or their beliefs. Charity is not an ethical embrace of cultural otherness but a recognition that the mind needs the world to furnish its content (Davidson 1984: 137). Descartes' *internalist* claim that we could be locked in our minds by an Evil Demon (ditto: mad neuroscientists or godlike aliens from another reality) is incoherent (§1.4). Although any particular belief can turn out false, massive error would deprive us of the transactions with things that give our beliefs worldly purport (Davidson 2001b: 153).

As well as opposing scepticism, Davidson's account seems to furnish an argument against the possibility of radically alien minds: no alien conceptual scheme, it seems, could be so strange as to resist interpretation since, as in the case of global scepticism, such beings *would lack the rational and coherent environmental transactions that could qualify them as thinkers at all* (§§5.7, 8.1). So, again, Vinge's tremulous speculations about radical weirdness seem off the mark.

Davidson's semantics, then, entails that the grounds for adhering to semantic theories are exhausted by the public facts about use accessible to a radical interpreter, and thus that incompatible theories of meaning could be equally consistent with the same observational data about what folk say, when. But if that is right, what meaning can be attached to the "shared world" about which speakers compare and contrast beliefs?

The obvious and intuitive answer is to respond that the shared world is just all the things there are. However, this universe of things would be raw material for interpreting a worldview via the incorporation of its members into different models. Both thinkers accept that there is no pragmatic way of ruling out the construction of multiple theories in line with Putnam's model-theoretic argument (see Davidson 1984: 235, 239–40). There could be (if we accept Putnam's metaphysical assumptions) no uniquely true theory related to it and thus we can make no sense of the idea that the unique, shared world is the collection of things in it.

Perhaps we might hope to live with this – accepting that the world is a determinate collection of lumps interpretable according to different lights. But this assumes without justification that the second-order vocabulary of terms like "object", "number" or "set" is uniquely assigned somehow even if first-order terms like "cat" and "cherry" are subject to disparate assignments. But which practices are supposed to "fix" the meanings of terms like "object"? As Putnam points out:

> "Object" itself has many uses, and as we creatively invent new uses of words, we find that we can speak of "objects" that were not 'values of any variable' in any language we previous spoke. (The invention of "set theory" by Cantor is a good example of this.)
>
> (Putnam 1988: 120)

Thus the indeterminacy that arises from the pragmatist conception of meaning applies recursively to the understanding of worldhood. Saying that the world is a model or a collection of lumps does not, as we had hoped, fix the contours of the real in a way that could explicate the idea of "world" presupposed by the possibility of radical interpretation and discursive subjectivity.

The claim that our shared world is *not* a lump presents obvious difficulties. First, radical interpretation means determining things like cats and rocks that form the topics of beliefs and statements. Thus this non-thing-like world must be compatible with the existence of determinable objects while not *being* those objects.

Such a world must also be unique in order to be shared. This is because, according to pragmatist theories of meaning such as Davidson's, agents qualify as believers by virtue of publicly interpretable transactions with a shared environment. If there are ghosts or intelligent cosmic dust clouds, then their actions might leave few detectable traces that humans can attune to. But such traces would have to be *interpretable in principle* by any intelligent interpreter given unbounded resources or time (see §§5.7, 8.1).

Thus a shared world in which the discursive subject operates must contain determinables and interpretables (texts) and it must be unique.

What could the shared world be if *not* a lump? Well, one plausible proposal derives from the holism thesis according to which every utterance or belief content is fixed by its relations to all the others (see above). These relations cannot be given in the way that the world according to MR is given. They are potentially infinite and also subject to differing but equally valid interpretations.

Some readers of Davidson – notably Jeff Malpas (1999) and Bjorn Ramberg (1989) – have employed the phenomenological idea of a horizon to explicate the idea of the world that underlies this pragmatic, interpretation-based account of meaning and mind.

As we saw, phenomenology is concerned with things as appearances and the conditions of that appearing. One general structure of givenness acknowledged by all phenomenological traditions is that a thing appears in a "wider structure of possible appearing" (Malpas 1999: 266). For example, Husserl claims that any perception of a thing is partial. If I see a hammer, I see it from a certain viewpoint, or hear it falling off a workbench as the cat passes by. If I think of it, I may represent it as a force amplifier or a birthday present. However, each thought or experience implies the possibility of nonactual perspectives. The hammer cannot be reduced to any of these: it is not determinate but, rather, *determinable* since its objectivity consists of being always *in excess* of its appearances (Mooney 1999). According to Husserl, this opening up of transcendence

is made possible by the complex nature of temporal presence, which always carries the anticipatory horizon of a new "now" (we will consider this account critically in §4.2).

In *Donald Davidson and the Mirror of Meaning*, Malpas argues that interpretation must have this horizontal structure. All interpretation occurs in a context fixed by certain interests and projects. Any particular project can be frustrated or break down (Malpas 1992: 128). Any project must, moreover, open onto the constitution of a new project, just as each view of the hammer implies the possibility of other views. Thus it is a normative assumption of this "interpretationist" position that each project of understanding is "nested" within further possible projects which extend to the totality of the psychological at an ideal limit.

This normative interleaving of interpretative projects is correlatively an interleaving of *things*. As we have seen, the pragmatist account of meaning cannot easily make sense of a *uniquely determinate* world with determinate representations of all the things in it. Beliefs cannot be identified independently of the determinables that believers interact with. By the same token, the identification of salient collections of objects and events occurs against the background of the interpreter's experience and interests. The nested structure of projects described by Malpas thus constitutes a plausible candidate for a non-reified "world" – a world not of things, but of potential "correlations" between intentional agents and determinable objects.

However, this interleaving is only intelligible if we assume each project to have a hermeneutic structure referred to as "fore-having" within the hermeneutic tradition. Each interpretation must potentially fan out onto future revisionary interpretations (Caputo 1984: 158). Without an appeal to anticipatory structure, there is little content that can be given to the idea of a single intersubjective world that Davidson and the other pragmatic-interpretationists must appeal to.

It is precisely at this point, according to Malpas, that the static concepts of the world (such as model theory) seem wholly inadequate and the *temporalized* models of intentionality and understanding developed in the phenomenological/hermeneutic tradition – represented by Husserl, Heidegger and Gadamer – assume importance.

Like Kant's philosophy, Husserl's mature phenomenology is a transcendental philosophy that supposes that any description of the world is a perspective of a transcendental subject that constitutes the sense of the world as an object of intentional experience (Mohanty 1989: 153). In his later writings, Husserl argues that our scientific and metaphysical theorizing is grounded in an unbreakable correlation between subjectivity and a perceptual "life-world" in which objects manifest themselves in relationship to our embodied activity. The various world concepts that have emerged in the history of science all depart from this life-world, but tend

to overlay it with metaphysical interpretations that estrange us from it. For example, Descartes and Galileo's separation of sensations like touch or colour from the allegedly "objective" geometrical properties of things seems like a natural assumption given the history of modern physics.

Husserl argues that this obviousness is a historical artefact which obscures the fact that the ideal shapes of geometry are not evident in our experience of perceptual objects. Whereas geometrical shapes are precise and distinct, the precision of ordinary objects is contingent upon our practical interests and technologies. The relation of geometrical abstractions to the life-world can, thus, be understood *historically* in terms of the practical privilege accorded linear or planar forms and the extension of this perfecting process in the thought of *ideal geometrical objects* (Husserl 1970: 26).

The Husserlian life-world is thus not a collection of lumps (a Putnam model) but something more like a "text" or field of determinables: a meaning or sense of an intentional experience (Rorty 1985).

Heidegger extends Husserl's accounts of time and the lifeworld by considering our non-cognitive relations to things, creatures and other people. He accepts that we represent the world as containing objective "present-at-hand" (*vorhanden*) things with bundles of properties. As we saw in Chapter 2, though, he also holds a version of the pragmatist claim that our cognitive access to the world depends on our noncognitive comportment in it (Okrent 2006). Heidegger conceives *Dasein* (the human agent) as essentially related to an environment of ready-to-hand (*zuhanden*) *equipment* which is practically "articulated" prior to any representation of their objective properties (Heidegger 1962: 98, 189; Okrent 2006). Thus in our regular dealings with it, the hammer is given in terms of its function, not represented as a bearer of objective properties that it could have independently of its function. A hammer is *something for* "producing, repairing or improving something". A car is *something for* transporting something (Heidegger 1995: 214). Novel applications – like using a car as a virility symbol (or murder weapon) – occur against the background of social norms prescribing the proper use of equipment (Okrent 2006).[11]

Equipment norms exhibit a *practical holism* that mirrors the psychological and semantic holisms in pragmatist theories of meaning. Hammers and nails belong to networks of functionally interrelated entities such as workbenches, planks and fences (Heidegger 1962: 120; 1995, 214; Dreyfus 1990: 62–3). Heidegger refers to these as "involvement totalities" (Heidegger 1962: 189; 1995: 215). We do not choose our involvement networks, according to Heidegger. They provide the context in which meaningful choices can be made and in which objects can come into view as salient topics of assertion. Each *Dasein* finds itself "thrown" into a socially meaningful context that it has not chosen but which allows certain

things and actions to show up as salient possibilities for it. However, thrownness requires that each *Dasein* is already absorbed into a background of shared concerns, activities and norms that *inter alia* grounds the normativity encountered in equipment (Heidegger 1962: 162).

Malpas argues that these structures furnish a "non-propositional horizon" against which entities "show up" as conforming or as failing to conform to our assertions about them. Meaning-holism and psychological holism reflect the structural openness of local constellations of practices to the things with which they are engaged. In particular, the presupposition of a common world which constitutes the horizon of radical interpretation is construed by him as inherent in the structure of disclosure itself:

> The appearing of something is the picking out (intending) of that thing from the wider structure (the horizon) of which it is a part. The wider network of possibilities is itself apparent (and then never completely) only when the project breaks down or is disrupted. In that disruption other possibilities come into view, even if only momentarily, as the project is reconstituted or as a new project arises. In that "moment" of truth the many possibilities of appearance (possibilities which can never be given complete specification) are freed up, only to be closed off again in the re-establishment of the project. What is shown in that moment, however, is the way in which things are within the horizon and the possibilities in which that horizon consists. Truth is the event of freeing up of possibilities which is also an opening up of possible appearances.
>
> (Malpas 1992: 257)

The practical-temporal structure outlined here unpacks the transcendentalist claim that subject and world are correlative rather than distinct. It also unpacks the active externalist view that the mental cannot be conceived other than as a unified pattern of activity on the part of situated, embodied agents (see also §2.4).

Looking forward

Our search for constraints on posthuman possibility did not yield much when confined to empirically refutable claims about the kinds of minds or information processing that might be possible in the physical world (§§2.1, 2.2). However, our search for potential *a priori* constraints seems initially promising. Two of the successors to Kantian transcendental humanism – pragmatism and phenomenology – seem to provide rich and plausible theories of meaning, subjectivity and objectivity which place clear constraints on 1) agency and 2) the relationship – or rather correlation – between mind and world.

These theories place severe anthropological bounds on posthuman weirdness for, whatever kinds of bodies or minds posthumans may have, they will have to be discursively situated agents practically engaged within a common life-world. In Chapter 4 I will consider this "anthropologically bounded posthumanism" critically and argue for a genuinely posthumanist or post-anthropocentric unbinding of SP.

Notes

1 If so, PPS = the empty set Ø.
2 This constraint obviously assumes the falsity of Cartesian dualism or similar doctrines (§1.4).
3 According to Sandberg, these limits are themselves far below the density of information that could be stored in the super-dense, degenerate matter found in collapsed stars.
4 Meaning that there is a universal Turing machine that could perform the operation given unbounded time and indefinite storage on its tape!
5 The disposition that we describe as the "sonorousness" of a metal tuning fork will be manifested differently in different atmospheric conditions. At sea level atmospheric pressure, the tuning fork that resonates at concert A will disturb the air around it and produce an audible pitch corresponding to the A above middle C on a piano (440 Hz). In a vacuum, there is no medium to disturb, thus no audible sound – though the fork will still resonate. However, the behaviour of the tuning fork will differ in each context because the friction caused by the air will "damp" its vibrations. Thus the harmonic properties and overall shape of the vibrations produced in the vacuum will be significantly different in each case.
6 Biological properties of nervous systems such as the noisiness of neuronal responses might be emulatable up to a point, but no further.
7 For example, functional differentiation in animal neural networks can be simulated without coding intracellular flows of ions for each "software neuron". This is because the learning processes that partition these networks into representational units depend on mechanism-independent principles such as the "Hebb rule" relating synaptic strength to the frequency of joint stimulation ("neurons that fire together, wire together").
8 Wennemann appends "human" and "posthuman" with a "B" superscript to indicate that "human" means *biologically human* rather than *morally human* and that "posthuman" entails *not biologically human*. This is because posthumans as agents would be *morally human* in his terminology.
9 Putnam makes the proviso that the interpretation function I must be "non-trivial" in that it assigns at least one predicate in the language F an extension (the set of things to which it applies) that is neither empty nor universal. If this holds we can use the *same universe of objects* over which I is defined to construct an "isomorphic" interpretation J in which every object in the extension of F is mapped one-one to some object belonging to F's different extension under J; thus the truth value of a sentence interpreted under I will be preserved under J while its meaning changes (Putnam 1981: 217). If F is universal and finite only an identity mapping can construct a set which maps one-one to the original extension to build the isomorphic interpretation by shuffling around or "permutating" objects from the same universe. This is trivially the case if F's extension is empty, since only the empty set has the same number of members as itself. However, non-triviality seems a fairly conservative assumption given that any minimally interesting theory about the world is liable to have some concepts which do not refer to everything or nothing. See Button (2013: 227–40) for a very well-illustrated discussion of the metamathematics of Putnam's "permutation argument".
10 That is, True or False.
11 The idea that environmental things are functionally typed in this way will be taken up again in my discussion of Jacques Ellul's philosophy of technology in §7.1.