On Reason and Spectral Machines: Robert Brandom and Bounded Posthumanism

David Roden

Bounded Posthumanism

Surveying the contemporary philosophical landscape, it seems that anthropocentrism remains relatively entrenched, if under stress. Post-Rawlsian political theory, for example, is largely concerned with distributing rights and entitlements to creatures supposed morally equal on the grounds of their humanity or personhood. The bioethics of technological enhancement and transhumanism is dominated by partially theorized assumptions about the moral priority of humanlike persons. Even left accelerationism, acutely sensitive to the ontological potential of drastic technoscientific change, pays lip service to anthropocentric notions of rational “self-mastery.”

However, anthropocentrism is not confined to practical philosophical sub-disciplines like ethics or political theory. Following Kant, modern transcendental thinkers propound a strong anti-naturalistic humanism for which humans are distinguished by the capacity to organize the world conceptually and semantically. According to these “transcendental humanists” the world is not structured independently of the concept-wielding activity that makes it an object of representation or practical agency.

A milder – that is, avowedly realist, or not ostensibly anti-realist – version of transcendental humanism is evident in the work of pragmatist readers of Kant such as Wilfred Sellars and Robert Brandom, for whom real intentionality and autonomy are ineluctably tied to the human capacity to follow public inferential proprieties. Brandom’s analytic pragmatism, as we shall see, draws a clear distinction between the ascribed intentionality of merely sentient beings, such as parrots, and the authentic intentionality of humans, whose capacity to institute and abide by shared norms rescue them from the pre-signifying state of nature.

Posthumanists are united by their rejection of philosophical anthropocentrism but tend to do so in different settings. In particular, critical posthumanism (CP) is concerned with the posthuman as a social and intellectual condition, while speculative posthumanism (SP) asserts the possibility of technologically made nonhuman agents. In other words, SP holds that there could be posthumans where posthumans would be “wide human descendants” of current humans that have become nonhuman in virtue of some process of technical alteration.

---

3 I’ve coined the term “wide descent” because exclusive consideration of biological descendants of humanity as candidates for posthumanity would be excessively restrictive. Posthuman-making technologies may involve discrete bio-technical modifications of the reproductive process such as human cloning, the introduction of transgenic or artificial genetic material or seemingly exotic processes like mind uploading. Thus entities warranting our concern with the posthuman could emerge via modified biological descent, recursive extension of AI technologies (including human and/or non-human designers), quasi-biological descent from synthetic organisms, a convergence of the above, or via some technogenetic process yet to be envisaged! (Roden 2012: 2014: 22)
5 This formulation allows that posthumans could be descended from technological assemblages that are existentially dependent on servicing “narrow” human goals. Becoming nonhuman in this sense is not a matter of losing a human essence
Nonetheless, CP and SP have a convergent interest in refuting a transcendental humanism that casts us as demigods, uniquely equipped to interpret a mute, unmeaning nature. Human life may be distinctive in many ways; but nothing obviously precludes comparable ontological novelty (arising, perhaps, from a bio-technological convergence) in the near or deep future. Accordingly, future life may diversify into alien forms quite different from those “our” cultural and evolutionary history equips us to understand. Exploring the implications of this “unboundedly weird” speculative posthumanism exposes the metaphysical commitments of anthropocentrism and thereby equips us to confront the ethical dilemmas posed by long-term technoscientific change.6

In Posthuman Life, I frame the tension between anthropocentricism and speculative posthumanism by distinguishing two claims about notional successors to current humans: an anthropologically bounded posthumanism (ABP); and an anthropologically unbounded posthumanism (AUP).

ABP is a strong corollary of transcendental humanism since it requires future agents to satisfy the conditions of current human agency. This position can be spelt out as follows:

1. There are unique constraints, C’s, [on cognition and agency] that all agents satisfy.
2. Any agent that knows it is an agent can correctly infer that the C’s applies to all agents (The C are transcendental conditions).
3. Human agents know they are agents.
4. Human agents can correctly infer that all agents satisfy the C’s.
5. Posthumans (if such existed) would be agents.7

_Human agents can correctly infer that posthuman agents would satisfy the C’s._

ABP’s purport becomes clearer if we consider the collection of histories whereby posthuman wide descendants of humans could feasibly emerge. In Posthuman Life I refer to this set as Posthuman Possibility Space (PPS).8 Given that posthumans would be agents of _some_ kind (See chapter 6 in Posthuman Life) and given ABP, members of PPS would have to satisfy the same transcendental conditions (C’s) on agency as humans.

Daryl Wennemann assumes something along these lines in his book Posthuman Personhood. He adopts the 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.”9 This is a condition of posthuman agency as much as of human agency. This implies that, whatever the future throws up, posthuman agents will be social and, arguably linguistic beings, even if they are robots or computers, have strange bodies, or even stranger habits. If so, PPS cannot contain non-anthropomorphic entities whose agency is significantly nonhuman in nature. ABP thus implies _a priori_ limits on posthuman weirdness.

---

6 See Roden, Posthuman Life, chapter 7 and 8.
7 See Roden, Posthuman Life, chapter 5 and 6.
8 Roden, Posthuman Life, 53.
AUP, by contrast, leaves the nature of posthuman agency to be settled *empirically* (or technologically). Posthumans might be social, discursive creatures; or they might be different from us in ways that we cannot envisage short of making some posthumans or becoming posthuman ourselves. AUP thus extends the critical posthumanist rejection of anthropocentrism to the deep time of the technological future.

In *Posthuman Life* I defended AUP via a critique of Donald Davidson’s work on intentionality; coupling this with a “naturalistic deconstruction” of transcendental phenomenology in its Husserlian and Heideggerian forms. Some of these arguments, I believe, carry over to Brandom’s overtly normativist philosophy.

The account of the relationship between normativity, social practice, intentionality that Brandom offers in *Making It Explicit*, and in other writings, is one of the most impressively detailed, systematic and historically self-aware attempts to explain subjectivity, agency and intentionality in terms of social practices and statuses. Moreover, contemporary proponents of philosophical realism and naturalism have praised Brandom for purging the Kantian tradition of anti-naturalistic, idealist and subjectivist deviations. If this is right, then Brandom’s views merit appraisal by posthumanists since they represent a comprehensive and powerful argument in favour of privileging anthropoform subjectivity and agency in the face of their speculative assaults on human-centered thinking. This is what I will attempt to undertake here, at least in a preliminary form. My aim is to show that despite the rigor and constructive brilliance of Brandom’s thinking, his thought fails to circumscribe notions of agency or meaning that could thwart the fissiparous decenterings of the human undertaken by SP and CP.

**First and Second-Class Agents**

I will begin with a thumbnail sketch of how Brandom derives conditions of possibility for agency and meaning from a theory of social practices. Then I will consider whether its foundations are capable of supporting this transcendental superstructure.

Brandom is a philosophical pragmatist. He claims that our conceptual and intellectual powers are grounded in our practical abilities rather than in relations between mental entities and what they represent. His pragmatism implies a species of interpretationism with regard to intentional content. Interpretationists, such as Daniel Dennett, claim that intentional notions such as “belief” do not track inner vehicles of content but help us assess patterns of rational activity on the part of other “intentional systems.” Belief-desire talk is not a folk psychological “theory” about internal states, but a social “craft” for evaluating and predicting other rational agents.

For Dennett, an entity qualifies as an agent with reasons if predicting its behaviour requires interpreters to attribute to it the beliefs and desires it ought to have given its nature and environment. A being whose behaviour is “voluminously predictable” under this “intentional

---


12 Brandom also follows Kant in trying to understand semantic notions like reference and truth in terms of their roles in articulating judgements rather than as semantic or representational primitives (Brandom, *Making It Explicit: Reasoning, Representing, and Discursive Commitment* (Cambridge, MA: Harvard University Press, 1994), 79-80).

stance" is called an “intentional system” (IS). In IS theory, there is no gap between predictability under the intentional stance and having real intentionality.\footnote{Daniel C. Dennett, The Intentional Stance (Cambridge, MA: MIT Press): 13-42.} \footnote{Intentional systems are unlikely to contain just sawdust or stuffing, but IS theory is agnostic regarding their internal machinery. Thus it undercuts both eliminativism and reductionism while providing a workable methodology for investigations into the mechanisms that actuate intentional systems.}

Brandom endorses Dennett’s claim that intentional concepts are fundamentally about rendering agency intelligible in the light of reasons, but argues that IS theory furnishes an incomplete account or intentionality. Interpretation is, after all, an intentional act; thus interpretationists need to elucidate the relationship between attributed intentionality and \textit{attributing} intentionality. If we do not understand what counts as a prospective interpreter, we cannot claim to have understood what it is to attribute intentionality in the first place.\footnote{Brandom, Making it Explicit, 59.}

Brandom goes one further. The intentionality attributed to intrinsically meaningless events or linguistic inscriptions seems entirely derived from interpreters. Similarly, with relatively simple IS’s. Maze-running robots or fly-catching frogs can properly be understood from the intentional stance – making them true-believers by Dennett’s lights. But their intentionality seems likewise observer-relative; \textit{derived} from attitudes of \textit{interpreting} IS’s.\footnote{Brandom, Making It Explicit, 60.} \footnote{Brandom, Making It Explicit, 60, 276.} To hold otherwise, he argues, is to risk a disabling regress. For if intentionality is derivative all the way up, there can be no real intentional attributions and thus no derivative (non-observer relative) intentionality.\footnote{Brandom, Making It Explicit, 60.} \footnote{Brandom, Making It Explicit, 61.}

Brandom claims that his theory can be read as an account of the conditions an organism must satisfy to qualify an \textit{interpreting intentional system}; that is, to warrant attributions of \textit{non-derived} intentionality rather than the as-if intentionality we can attribute to simpler organisms or complex devices:

The theory developed in this work can be thought of as an account of the stance of attributing original intentionality. It offers an answer to the question, What features must one’s interpretation of a community exhibit in order to be said to be an interpretation of them as engaging in practices sufficient to confer genuinely propositional content on the performances, statuses, attitudes, and expressions caught up in those practices?\footnote{“The key to the account is that an interpretation of this sort must interpret community members as taking or treating each other in practice as adopting intentionally contentful commitments and other normative statuses” (Brandom, Making It Explicit, 61).}

Whatever else the capacity for original or “first class” intentionality includes, it must involve the ability to evaluate the cognizance and rationality of their actions and the actions of other beings.\footnote{Davidson, “Thought and Talk”, in Inquiries into Truth and Interpretation (Oxford: Clarendon Press, 1984): 155-170.} Entities with the capacity to assess and answer to reasons in this way are referred to by Brandom as \textit{sapient}. Entities with only derived intentionality may exhibit the \textit{sentient} capacity to react in discriminating and optimizing ways to their environment, but the conceptual content of these responses is attributed and observer-relative.

The claim that intentionality (or the capacity for objective thought) implies the capacity to evaluate reasons obviously has a rich post-Kantian lineage. However, one of the clearest arguments for connecting intentionality and the capacity for other-evaluation is provided by Donald Davidson in his essay “Thought and Talk.” Davidson begins with the assumption that belief is an attitude of “holding” true some proposition: for example, that there is a cat behind
that wall. If belief is holding true it entails a grasp of truth and the possibility of being mistaken; and thus a concept of belief itself. We cannot have a concept of belief without exercising it. Thus we cannot believe anything without the capacity to attribute to others true or false beliefs about the same topic.

This capacity presupposes linguistic abilities, according to Davidson, because attributing contents to fellow creatures requires a common idiom of expression. Absent this, the possession of a concept of belief and, thus, the very *having of beliefs*, is impossible.

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 intentionality 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.

Brandom agrees! We need language to have and attribute beliefs, and, by extension, practical attitudes corresponding to desires and intentions. However, his official account avoids talk of beliefs or intentions in order to steer clear of the picture of beliefs, etc. as “inner” vehicles of content (sentences in the head, say) rather than social statuses available to discursive creatures like ourselves.

For Brandom, the primary bearers of propositional content are public assertions. Thus he bases his elaborate theory of intentionality not on a theory of mental representations or sub-propositional concepts, but on a pragmatic account of the place of assertions within the social game of “giving and asking for reasons.” Correlatively, Brandom’s semantics begins with an explanation of how assertions — and their syntactical proxies, sentences — acquire propositional content. Like Sellars’ brand of functional semantics, it is framed in terms of the normative role of utterances within social practices that determine how a speaker can move from one position in the language-game to another (language-transition rules), assume an “initial position” (language-entry rules) or exit the game.

In the case of assertions, the language-transition rules correspond to materially correct inferences such as the inference that *x is coloured* from *x is red*. Language-entry rules are non-inferential since they are made on the basis of reliable dispositions to discriminate the world in inferentially consequential ways. As Brandom puts it, statements like “This is red” (uttered in response to red things) are “noninferentially elicited but inferentially articulated.” Finally, “language exit rules” correspond to practical commitments disposing one to forms of non-linguistic action.

Thus Brandom agrees with other post-Wittgensteinian pragmatists that linguistic practices are governed by public norms, as well as by reliable differential responsive

---

27 His subsequent, very rich, analysis of subsentential expressions is necessarily decompositional rather than compositional (Brandom, *Making It Explicit*, 79-82). For example, the difference between a predicate and a singular term is understood in terms of the different inferential consequence which follow from their inter-substitution within sentences (368).
dispositions (RDRD’s). However, he follows Davidson in rejecting the “I-we” conception of social structure.\textsuperscript{31} If meanings are inferential roles (as Dummett and Sellars also claim), then the content attributable to expressions will dance in line with the doxastic commitments of individual speakers.

Suppose one observes a masked figure in a red costume clambering up a skyscraper. The language entry rules you have internalized may entitle you (by default) to claim that Spiderman is climbing the building. However, you are unaware that Spiderman is none other than Peter Parker. So you are not yet entitled to infer that Peter Parker is climbing the building – although the “substitution-inferential” rules of English would entitle you to that further claim if (say) some reliable authority informed you of this fact.\textsuperscript{32}

This simple example shows that the inferential roles – thus meanings – of expressions like “Spiderman” are not fixed communally but vary with auxiliary assumptions, sensitivities and dispositions of individual speakers. Understanding or interpreting the utterances and beliefs of others is a matter of deontic scorekeeping – that is, keeping track of the way social statuses alter as speakers update their inferential commitments.\textsuperscript{33}\textsuperscript{34} Thus semantic and intentional content are co-extensive with the normative-functional roles of states and actions. It follows that what a belief or claim “represents” or is “about” is fixed by the status it can be ascribed from the perspective of various deontic scorekeepers (including the believer or claimant).

Functional semantics can be thought of as a philosophical appropriation of the formal conception of computation as automated symbol manipulation developed in the early part of the twentieth century.\textsuperscript{35} In its purely mathematical form a computational engine can be understood as a set of “state transitions” fixing how the data stored in some memory location determines consecutive states of the machine. A Turing machine’s table, for example, “completely determines” how it would behave when reading a particular symbol at a particular memory location on its “tape” while in a particular state.\textsuperscript{36}\textsuperscript{37}

The obvious attraction of this socio-mechanical metaphor to contemporary philosophical materialists like Ray Brassier is that it promises to cash out abstract notions such as “meaning” or “representation” in terms that are, at first sight, closer to home: people uttering and inscribing marks; responding and acting to the impulses of a shared natural world.\textsuperscript{38}

However, with Brandom’s social machines the inferential transitions are not formally or causally determined but required or permitted. As we shall see in the following two sections,


\textsuperscript{32} I.e. that the inferential move in the language game from “Spiderman is climbing the building” to “Peter Parker is climbing the building” is materially valid.

\textsuperscript{33} Brandom, \textit{Making It Explicit}, 142.

\textsuperscript{34} The point of attributions of belief or desire, for example, is to determine what an agent is committed or entitled “to say or do.” Likewise, the point of affixing truth values to beliefs or statements is to assess or endorse their propriety within the game of giving and asking for reasons. Is the claimant entitled to assert that p? Are the inferential consequences of p that they acknowledge the actual consequences? (17, 542).

\textsuperscript{35} Sellars, for example, is happy to accept that learning to infer, is at base, a matter of internalizing formal transformation rules. Sellars, “Some Reflections on Language Games,” 209.


\textsuperscript{37} The table specifies which operation the machine carries out when in a particular machine state (say, q0) and a particular symbol is lying on square current being scanned. The table may, for example, specify that if the machine is in q0 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. q2). 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, a the binary expression of a fraction (See Petzold 2008).

this network of proprieties (and thus the social apparatus they compose) is somewhat spectral and elusive since Brandom does not consider them to be factually real but spun from the passing attitudes of the scorekeepers.

Brandom, like Davidson, argues that the ascription and adoption of such states is only possible if the scorekeepers can practically express them in a structured language with components such as predicates and singular terms. Thus, as advertised, Brandom’s account suggests a pragmatic-semantic story with which to transcendentally partition PPS. If posthumans are to be intentional agents in thrall to concepts, they will be subjects of discourse assessing one another according to public inferential proprieties.

The Norm-Grounding Problem

However, we have grounds for partitioning PPS along these lines only if normativism can contend with some difficult foundational issues deriving from the aforementioned spectrality of inferential roles. I will refer to the most pressing of these as “the norm-grounding problem”.

Brandom’s pragmatics implies that the rules that furnish deontic statuses are implicit in what we do: in our linguistic and non-linguistic performances, rather than in some explicit set of semantic rules. But what does it mean for a norm to be implicit in a practice? What is it about what we do that constitutes our observance of one norm rather than another? Are norms a special kind of fact, to which our practices conform or fail to conform? If there were normative facts that transcended our actions, this could at least explain how our inferences can be held to account by them.

Brandom rejects factualism regarding norms. They are not, he claims, “part of the intrinsic nature of things, which is entirely indifferent to them.” This seems wise on the face of it. If there are Platonic norms, it is far from clear how animals like us, or our evolutionary forebears, could come to be aware of them. Brandom thus adopts a nonfactualist or “phenomenalist” position regarding norms. Non-normative reality is “clothed” in a weave of normative statuses when speakers treat public actions as correct or incorrect, permitted or entitled.

However, before considering Brandom’s nonfactualist account of norms in greater detail, it is instructive to consider a superficially appealing position that he rejects: regularism. Regularism is the claim that norms are regularities. To act according to a norm (or follow a rule) is simply to conform with a regularity. Regularism is consonant with pragmatism because one can conform to a regularity without one having explicit knowledge of it. This avoids the vicious regress that ensues if we require that semantical rules need to be explicitly grasped by speakers. Regularism also appeals to philosophical naturalists because it explains how norms depend (or supervene) on facts about the physical state and structure of individual speakers.

However, Brandom rejects this attempt to ground normative claims in factual claims. Here he follows Kripke’s seminal reading of Wittgenstein’s discussion of rule-following:

---

39 Brandom, Making it Explicit, chapter 6.
42 Brandom, Making It Implicit, 48.
43 Brandom, Making It Implicit, 27.
44 Brandom, Making It Implicit, 24-25.
pointing out that any finite sequence of actions will conform to a possibly infinite number of regularities. Thus suppose, as in Kripke’s original example, that all the addition sums I performed involve values less than 57. My addition behaviour is consistent with the function that always maps two values onto their sum. But it is also consistent with the function “quus” that maps two numbers onto their sum if each is less than 57 and onto 5 otherwise.\(^\text{45}\)

There will be an infinite number of such interpretations of my arithmetical practice, no less consistent with it than with the plus function. And this situation will apply for any maximum summed values.\(^\text{46}\) So any episode of my supposed “additive” behaviour will be equally interpretable as quaddative. Similar considerations apply for any maximum of summed values and extend to empirical concepts, as the easy definition of “gruesome” predicates illustrates. Thus historical applications of the term “horse” to its instances are behaviourally consonant with the rule for the predicate “shmorse” – which applies to a thing if it is observed before the year 30,000 and is a horse, or is not observed and is a cat.

The take home moral is that there is no such thing as the unique regularity that a finite performance conforms to. Moreover, for any continuation of that performance “there is some regularity with respect to which it counts as ‘going on in the same way’.”\(^\text{47}\) There are just too many ways of gerrymandering regularities for any given continuation of a performance; and the simple regularity view provides no basis for selecting between them. So the simple regularity account fails to explain how a determinate norm can be implicit in practice.

One appealing response to the failure of the simple regularity view is to shift attention from finite stretches of performance “to the sets of performances (for instance, applications of a concept) the individual is disposed to produce.”\(^\text{48}\) The appeal of unpacking the idea of grasping a rule in terms of dispositions is that it might be that one could be disposed to do an infinite number of things that one never gets round to doing due to lack of time or the absence of triggering input.\(^\text{49}\) So a dispositional analysis\(^\text{50}\) seems to get at the infinitary nature of concepts in naturally admissible ways.

So it appears that we can avoid the gerrymandering objection by saying that different agents A and B grasp the same rules where they are disposed to perform identically given the same triggering inputs. However, Brandom rejects the dispositionalist account of rule following. Following Kripke he claims that dispositionalism is unable to account for misapplications of a rule. Abiding by a rule has to be compatible with errors in performance.\(^\text{51}\) One can violate a norm. But one cannot, he claims, act in violation of one’s dispositions. For example, A might be disposed to behave identically under the same triggering conditions as B. But whereas A is correctly applying the adding rule, B could be incorrectly quadding or misapplying some other rule.\(^\text{52}\) If a dispositionalist grounding of norms treats dispositions counterfactually, then, it will be unable to account for the mismatch between the rule followed

---


\(^{46}\) So in addition to quus, we can define the function wuss. Where

\[
x \text{ wuss } y = \begin{cases} 
  x + y & \text{if } x, y < 58 \\
  5 & \text{else}
\end{cases}
\]

And so on!


\(^{49}\) So is not necessary for the rule user to have all the triggering instances “before his mind” to have grasped how to perform in any of these instances.


\(^{52}\) By extension a straight dispositional analysis would not distinguish someone who means grue by “green” but has been neurologically interfered with so that they would apply it (incorrectly) to green things after the year 2010 and a user of “green” who would be disposed to apply it to exactly the same items.

**Deontic Statuses and Deontic Attitudes**

As advertised, Brandom’s favoured account of norms is nonfactualist. We “clothe” a non-normative world in deontic statuses, he claims, by *taking* certain actions or utterances to be correct or incorrect.\footnote{In “Rules and Powers,” Martin and Heil present a very interesting case for holding that dispositional accounts of rule following can avoid Kripkensteinian skeptical conclusions if dispositions are construed realistically rather than in terms of statements about counterfactual behaviour. Then it could be true of A and B that they would perform identically even though A is disposed to follow the plus rule whereas B is disposed to follow the minus rule. This is because B’s disposition could be “blocked” in some way accounting for the error. Howhy 2003 develops a similar account.} Normative statuses arise only insofar as there are creatures that can treat *one another as committed or entitled to do this or that*. In Brandom’s terminology: deontic statuses are assigned when creatures adopt *deontic attitudes* towards one another.

Looking at the practices a little more closely involves cashing out the talk of deontic statuses by translating it into talk of deontic attitudes. Practitioners take or treat themselves and others as having various commitments and entitlements. They keep score on deontic statuses by attributing those statuses to others and undertaking them themselves. The significance of a performance is the difference it makes in the deontic score – that is, the way in which it changes what commitments and entitlements the practitioners, including the performer, attribute to each other and acquire, acknowledge, or undertake themselves.\footnote{Brandom, Making It Explicit, 161.}

But what are deontic attitudes? If – like *propositional* attitudes – they are inherently intentional Brandom is stuck in a regress. The philosophical attraction of normative functionalism is that it promises to reduce intention-talk to norm-talk. If deontic attitudes are necessarily intentional, however, he has made little progress in explaining interpreting intentionality in terms of social practices. Moreover, his account would fail to accord with a modest Darwinian naturalism regarding the emergence of the intentional. The requirement I have in mind is Darwinian in the loose sense that it holds that the intentional and the mental are not basic features of the world but depend on the way heterogeneous arrangements of mindless (or less minded) things interact with one another.\footnote{Levi R. Bryant provides a useful summary of how this kind of naturalism can be cashed as a Deleuzian Machine-Oriented Ontology (MOO) in “The Gravity of Things: An Introduction to Onto-Cartography”, *Anarchist Developments in Cultural Studies* 2 (2013): 10-30.} Note that this commitment is not restricted to analytic naturalists like Dennett or Jerry Fodor. It applies to continental materialists and critical posthumanists who, in Pramod Nayar’s words seek “the radical decentering of the traditional sovereign, coherent and autonomous human in order to demonstrate how the human is always already evolving with, constituted by and constitutive of multiple forms of life and machines.”\footnote{Nayar, Pramod K. *Posthumanism*. Cambridge: Polity, 2013.} They have to deny the existence of basic psychological properties for such a project to have any prospect of decentering *anything*. The decentering effect of the claim that humans are constituted by congeries of machines and life-forms presupposes that these

\footnotesize
54 In “Rules and Powers,” Martin and Heil present a very interesting case for holding that dispositional accounts of rule following can avoid Kripkensteinian skeptical conclusions if dispositions are construed realistically rather than in terms of statements about counterfactual behaviour. Then it could be true of A and B that they would perform identically even though A is disposed to follow the plus rule whereas B is disposed to follow the minus rule. This is because B’s disposition could be “blocked” in some way accounting for the error. Howhy 2003 develops a similar account.
55 Brandom, Making It Explicit, 161.
56 Brandom, Making It Explicit, 166.
57 Levi R. Bryant provides a useful summary of how this kind of naturalism can be cashed as a Deleuzian Machine-Oriented Ontology (MOO) in “The Gravity of Things: An Introduction to Onto-Cartography”, *Anarchist Developments in Cultural Studies* 2 (2013): 10-30.
heterogeneous entities are not themselves the kind of thing that possess autonomy or subjectivity as classically conceived.

However, such positions need to go further and provide explanations for the emergence of mental properties in a non-mental world or (if they are deflationary) for the emergence of systems that can attribute such psychological states. This is because they must deny that such powers emerge “spookily” – i.e. in ways that are recalcitrant to explanation. This point is recognized, for example, by the Deleuzian philosopher of science Manuel DeLanda who proposes that any explanation of emergent behaviour in a given system should have a mechanistic component framed in terms of its constituents and the ecological relations they enter into: for example, a system of chemical reactants far from equilibrium, or a population of individuals in a pre-state society. Spooky or strong emergence would de-fang the decentering effect by allowing subjectivity to jump fully formed out of the slime of heterogeneity even where these attributes do not form part of the basic furniture of the world.

Naturalists, materialists and posthumanists should, then, require that our theories of intentionality be compatible with some gradualist explanation of the development of intentional systems from non-intentional ones. In this instance, that norm-instituting powers cannot have appeared fully formed but must have emerged gradually from the scum of sentience.

Brandom is properly sensitive to these requirements. As he puts: “It is clear that there were nonlinguistic animals before there were linguistic ones, and the latter did not arise by magic.” The capacity to ascribe deontic and practical commitments in discourse presupposes a story whereby “suitably social creatures can learn to distinguish in their practice between performances that are treated as correct by their fellows”. Darwinian naturalism thus enjoins Brandom to show how deontic attitudes can occur in “pre-linguistic communities” that lack full noetic and agential powers.

The simplest model of deontic attribution that he provides is one in which performances are assessed as something the performer is authorized to do by the withholding of sanctions – where sanctioning behaviour, here, is “compounded out of reliable dispositions to respond differentially to linguistic and nonlinguistic stimuli” not florid interpretative powers. For example, the deontic status of being entitled to pass through a door might be instituted by a ticketing system in which “the ticket-taker is the attributer of authority, the one who recognizes or acknowledges it and who by taking the ticket as authorizing, makes it authorizing, so instituting the entitlement.” This account can be complicated if we introduce deontic attitudes that institute responsibilities on the part of agents. For example, taking the Queen’s shilling makes one liable to court martial if certain military duties are not undertaken.

According to Brandom these cases illustrate how social actors can partition “the space of possible performances into those that have been authorized and those that have not, by being disposed to respond differently in the two cases.” Does this model show that Brandom’s account can satisfy the minimal naturalist constraints that he recognizes?

60 In DeLanda’s account the second component corresponds to the Deleuzian Idea: the specification of singularities reflecting that same system’s tendency to slip into distinctive portions of its state space.
61 Rosen, “Who Makes the Rules Around Here?”
64 Brandom, *Making It Explicit*, 156.
A number of commentators – including Daniel Dennett and Anandi Hattiangadi – have argued that it succumbs to the same gerrymandering objections that Brandom cites against regularism. If so, any performative regularities (actual or counterfactual) exhibited by actors and sanctioners in this simple model will be consistent with multiple normative readings of either behaviours – including interpretations that render the “deontic attitudes” mistaken.

The response of a ticket-taker towards ticket holders for some prelinguistic social event might be to open a door, physically permitting entry. However, this is consistent with multiple deontic statuses: including “entry permitted unconditionally,” “entry permitted [mistakenly] according to a rule barring ticket holders and nonholders alike,” “entry permitted, conditional upon possession of a ticket and having a birth place within a five-mile radius” and so on. Thus if deontic attitudes have to be compounded only from reliable dispositions to respond to stimuli, each such response is consistent with the attribution of many, many mutually-exclusive statuses.

Just as there is nothing corresponding to the regularity exhibited by a given stretch of sign-using behaviour, there is nothing corresponding to the deontic status exhibited by any finite episode of what (at first sight) appears to be sanctioning behaviour. Otherwise put, if regularism fails to supply the constitutive ground for meanings (as Brandom claims) it fails just as spectacularly to supply the constitutive ground for deontic status ascriptions.

As Hattiangadi points out, beefing up the mental powers of instituters would avail Brandom little, or rather too much. If we furnish sanctioners with the power to make contentful judgements (about whether an agent is entitled to pass through the door, for example) we are already in the realm of the intentional. This indeterminacy ramifies equally if we suppose the sanctioning behaviour extended to something resembling sign use. Suppose pre-linguistic Emma sanctions pre-linguistic John by kicking him when the latter points to something saying “That’s red”.

The question is what has John been punished for? Has Emma attributed the commitment to say ‘that’s not blue’, or has she attributed the commitment to say ‘that's not grue’? Which of these commitments has John violated?

Again, we cannot attribute contentful attitudes regarding the regularity that John failed to follow here without attributing the florid intentional and agential powers whose emergence was to be accounted for by their prelinguistic analogs. If regulism is false, reliable responses alone do not suffice to furnish contentful attitudes regarding correctness or incorrectness of others’ performances. Thus – prior to the emergence of sapience – there can be no deontic statuses at all.

It follows that a naturalistically constrained normativism does not appear able to explain how social but nonlinguistic beings can institute norms, thus normative statuses, without a vitiating appeal to florid intentional powers. But this explanatory gap implies that Brandom cannot provide an explanatory framework in which the emergence of intentionality and sapience are non-magical.

The Interpretationist Defence

Can Brandom’s account be repaired in a way that meets his minimal naturalist commitment?

Well, one defence that seems consistent with Brandom’s avowals elsewhere is to follow Davidson and Dennett by claiming that the certain kinds of social behaviour are norm-governed if a) members of our speech community would properly interpret them as normative or b) if an ideally rational interpreter privy to all the relevant behavioural facts would read them as normative. This response has something to recommend it. When interpreting alien social practices, we are liable to appeal to our own background assumptions about what performances belong to the kind “social practice.” Moreover, appealing to notion of an ideal interpreter can be of value when trying to understand the theoretical and empirical constraints on attributions of semantic or normative content.

However, as Hattiangadi remarks, this response misses the point of the dispositional analysis of deontic attitudes. This was to explain how a non-sapient community could bootstrap itself into sapience by setting up a basic deontic scorekeeping system.72 Appealing to actual or ideal interpreters simply replicates the problem with Dennett’s intentional stance approach since it tells us nothing about the conditions under which a creature qualifies as a potential interpreter and thus little about the conditions for meaning, understanding or agency.

A similar problem afflicts Joseph Heath’s proposal73 that Brandomian norms emerge from reciprocal expectations supported by sanctions. The idea is that a first person acts in a certain way while expecting a sanctioning response from a second person. The second person, meanwhile, is disposed to respond to certain performances with sanctioning behaviour while the first person recognizes this. Where this minimal intersubjective couple converges towards a single pattern of behaviour over time, Heath argues, we are entitled to treat their activity as implying a norm.

Heath’s proposal may be fine if we assume that certain intentional powers are already in place – e.g. that each individual both expects and sanctions the activity of the other. However, as Hattiangadi’s appeal to the gerrymandering argument shows, this structure presupposes beings capable of intentional states such as expecting and sanctioning. This is presumably what distinguishes it from simpler cases of dynamical coupling where two physical systems converge towards a single pattern of behaviour. But if the normativist is serious about explaining the intentional in normative terms, they are not entitled to these assumptions.

**Unbounded Posthumanism**

If Brandom is right about the defects of Dennett-style or Davidson-style interpretationism, the tendency for his own account to regress to those positions is most telling. It suggests that interpretationist accounts cannot explain the semantic or the intentional without regressing to assumptions about ideal interpreters or background practices whose scope they are incapable of delimiting: “[In] principle interpretability is ill defined unless we have some conception of what is doing the interpreting.”74

The point is not that interpretationism is false but that it is unilluminating. It is empirically unproblematic that we interpret other speakers, texts, cultural artifacts, etc. However, if in-principle interpretation according to the intentional stance fixes the content of intentional discourse, but the nature of such interpretation is ill defined, we have merely satisfied our curiosity about the nature of mindedness by appealing to local mind-reading techniques. We do not yet know what the invariants (if any) of intentional interpretation are.

---

Another way of putting this is that our practices of interpretation and deontic assessment are phenomenologically “dark.” The fact that we have them and have a little empirical knowledge of them leaves us ignorant both of their underlying nature and (by extension) of the space of interpretative and psychological possibility. Normativist ABP and its interpretationist variants thus provide no future-proof constraints on the space of possible minds or possible agents. Anthropologically Unbounded Posthumanism is not seriously challenged by the argument that mind and meaning are constituted by social practices. AUP implies that we can infer no claims about the denizens of Posthuman Possibility Space a priori, by reflecting on the pragmatic transcendental conditions for semantic content. We thus have no reason to suppose that posthuman agents would have to be subjects of discourse or members of communities.

For example, it is conceivable that there might be beings that are far more capable of altering their physical and functional structure than current humans. I call an agent “hyperplastic” if it can make arbitrarily fine changes to its structure without compromising its agency or its capacity for hyperplasticity. A modest anti-reductionist materialism of the kind embraced by Davidson and (arguably) Brandom implies that such agents would be uninterpretable using an intentional idiom because intentional discourse could have no utility for agents who must predict the effects of arbitrarily fine-grained self-interventions upon future activity. However, the stricture on auto-interpretation would equally apply to hetero-interpretation. Were such hyperplastics possible, they would not be interpretable for discursive creatures, which is not to say that they would be uninterpretable tout court.

As Scott Bakker and I have argued, this position is fatal for the ambitious rationalist or “Promethean” projects of thinkers such as Brassier and Reza Negarestani. These are inhumanist insofar as they reject the claim that a commitment to Enlightenment entails a commitment to any ontological or theological conception of the human subject. Inhumanism proposes that all meaningful intelligence is artificial insofar as it involves the unbounded extension of discursive practices: humanity just is the revisionary power to redefine humanity within the discursive space of reasons. However, AUP implies that there is no warrant for the claim that any serious intelligence must be a “subject of discourse” able to measure its performances against public standards. So the space of possible intelligences and agents is notionally far larger and stranger than can be accommodated by Brassier and Negarestani’s bounded inhumanism.

By extension, the politics of posthumanism cannot be fixed by the structure of discursive agency either. We have no future-proof grasp of how strange posthumans (our wide-descendants) might be, so we lack any basis for adjudicating the moral status of such beings. We may buy into a parochial humanism which accords human subjects a level of moral consideration that is greater than the nonhuman creatures we know about. But this does not entail that there are not ethically considerable states of being in PPS that have little in common with the modes of being accessible to current humans. If posthuman politics is

---

75 See Roden, “Nature’s Dark Domain”; Posthuman Life, 82-104.
anthropologically unbounded, in this way, any ethical assessment of the posthuman must follow upon its historical emergence. If we want to do serious posthuman ethics, we need to make posthumans, or become posthuman.

**Bibliography**


