

SPHEXISHNESS, EPISTEMIC BOUNDS, AND A PRIORI PSYCHOLOGY

M. P. Smith

Department of Computer Science

Duke University

We are all too aware of limitations on human cognitive capacities. All of us routinely run up against the boundaries of our abilities to calculate swiftly, to take several factors into account at once, to recall information. Many of the experimental techniques of cognitive psychology involve the measurement of these limits. Lately it has been suggested that we may expect the human mind to have limitations of a different order, qualitative rather than quantitative in character. These are not merely upper bounds on the size, speed, reliability, etc. of calculation, but limitations on the scope of human understanding. I have in mind here Fodor's claim that our minds are likely epistemically bounded and the Dennett-Hofstadter thesis that they are sphexish.

These claims are not without precedent. Fodor cites Aquinas, who borrows from Aristotle, who attributes to Anaxagoras the argument that, since the intellect can think all things, it must be "unmixed with body" (De Anima 111.4).¹ Dennett cites Descartes (Dennett 1978, p. 245). Traditionally the claim that materialism implies limitations on the understanding has been used to argue for immaterialism. This immaterialist modus tollens has become the materialist's modus ponens. In this paper I consider the justification for the implication itself, however it may be used, and also try to clarify the content of the claim being made.

Sphexishness has as its paradigm case the behavior of the sphex wasp as described in an oft-quoted passage from Wooldridge (Wooldridge 1963, p. 82). The wasp's behavior of leaving her prey at the entrance of the burrow, entering unencumbered at first, and then retrieving her burden, smacks of prudent foresight. One can't be too careful! But if, while she is checking things out inside, a meddlesome finger nudges her prey a few inches from the doorway, she will repeat the procedure, even though presumably she can see that no enemy has snuck into her burrow in the meantime. Further, she will repeat this routine endlessly, or at least to the end of an experimenter's patience. So one can be too careful. After a few dozen repetitions, however, it becomes obvious that the wasp is not being careful. At first she seems prudent, then stupidly over-cautious. Eventually, though, we abandon the whole framework of rational appraisal and realize that the poor wasp is just stuck like a broken record (Hofstadter 1985, p. 530). Sphexishness is apparently rational behavior revealed to be really merely reflexive through its rigid persistence where it is no longer appropriate. Hofstadter calls the sphex's performance "a rather shocking revelation of the mechanical underpinning ... of what looks like quite reflective behavior" (Hofstadter, p. 529).

Note that sphexish behavior is not merely sub-optimal or even irrational, but non-rational. Admittedly the line between behavior which is irrational and that which falls outside the pale of rational criticism is not sharp. To claim that some human behavior is sphexish, however, is to say that it is only apparently done for reasons, not that it is done for poor reasons.

Wooldridge's original analysis is that the rigidity of the sphex's apparently thoughtful behavior reveals it to be pre-packaged and triggered,

rather than deliberately chosen. He compares it to an interrupt-driven subroutine (Wooldridge, pp. 83 - 85). Sphexishness would arise in such a system when indiscriminating interrupts are in charge of extended behavioral routines.

At the end of Part V of the Discourse on the Method Descartes proposes two features which separate people from both animals and machines. The first and more famous is the ability to use language. The second has not been well understood, but I believe that it amounts to a claim that people are antisphexish.

About the abilities of animals Descartes says:

...although many animals show more skill than we do in some of their actions, yet the same animals show none at all in many others; so what they do better does not prove that they have any intelligence, for if it did then they would have more intelligence than any of us and would excel us in everything. It proves rather that they have no intelligence at all, and that it is nature which acts in them according to the disposition of their organs. (Cottingham et al. 1985, p. 141)

Clearly Descartes must be referring to apparently rational abilities on the part of animals, and not just, say, the cheetah's speed or the monkey's agility, for his denial that these skills prove intelligence to be apposite. Descartes is claiming that animal cunning is in fact just sphexishness.

Why Descartes thinks that the singularity of the spider's skill at construction or the bird's navigational prowess shows these to be merely mechanical rather than truly intelligent behaviors comes across more clearly when Descartes distinguishes between the abilities of people and machines.

...even though such machines might do some things as well as we do them, or perhaps even better, they would inevitably fail in others, which would reveal that they were acting not through understanding but only from the disposition of their organs. For whereas reason is a

universal instrument which can be used in all kinds of situations, these organs need some particular disposition for each particular action; hence it is for all practical purposes impossible for a machine to have enough different organs to make it act in all the contingencies of life in the way in which our reason makes us act. (Cottingham et al., p. 140)

Before the term 'sphexishness' was coined by Hofstadter, Daniel Dennett, citing Descartes, claimed that the lesson of the sphex was that "any behavior controlled by a finite mechanism must be tropistic" (Dennett 1978, p. 245). Apparently Dennett meant by a "finite mechanism" an elaboration of the sphexian architecture as imagined by Wooldridge: a system of hard-wired patterns of response set off by pre-arranged signals. The Dennett-Descartes claim then would be that no finite system of routines and interrupts could act with the resourcefulness of a being endowed with universal reason.

This thesis is indefensible, however, simply because a finite system can still be very large. The number of situations a human being can expect to meet, with however small a delta factor of probability one chooses, is still finite. (I say this while cheerfully admitting I have no idea how to count possible situations, much less estimate their probability. But the product of the number of nanoseconds in the maximum human life span times the number of cubic angstroms within, say, thirty parsecs of earth, times the number of distinct physical states each of these space-time parcels can be in, is still comfortably finite.) Finitude is no constraint at all when there are only finitely many opportunities. Dennett should have followed Descartes' lead and talked about the limitations of biologically or physically feasible mechanisms, rather than merely finite ones.

A system with a finite number of primitive responses is still capable of an infinite number of responses anyway, provided that we allow combinations of

responses of any length. The advantage of universal reason over finite mechanism is not to be determined by the cardinality of behavioral repertoires. If the alleged human sphexishness amounts to more than the admission of upper bounds on the size, speed, and accuracy of the information processing that goes on within them, then it must be based on more than human finitude.

II

It is tempting to read Descartes' argument for the sphexishness of animals as anticipating Fodor on the consequences of modularity. If intelligence consists in computationally discrete, informationally encapsulated, special-purpose organs, "then it is surely in the cards that there should be some problems whose structure the mind has no computational resources for coping with" (Fodor 1983, p. 120). In particular, Fodor proposes that a likely consequence of the mind's modularity is that it is epistemically bounded: "our cognitive organization imposes epistemically significant constraints on the beliefs that we can entertain" (ibid.). If there are beliefs that we cannot entertain, then there are reasons for acting that we cannot have. In situations where those reasons would be significant we should act sphexishly. Anticipating objections on this point, however, Fodor goes on to argue that epistemic boundedness is not unique to modularism, but in fact should be a consequence of any psychological theory. If Fodor is right, we can know a priori that we or any cognitive beings would be sphexish.

Fodor offers two sorts of reasons apart from modularity to expect the mind to be epistemically bounded. The first sort involves bounds on executional parameters: certain beliefs might be so complex that their mental tokens are unparseable, too large for short-term memory, or the like (p. 121). The more

interesting possibility is that the mind's basic conceptual repertoire might be limited. Empiricism -- nothing in the mind unless first in the senses -- poses such a restriction on the concepts available to the mind, and Fodor points out that empiricists like Hume are saved from epistemic boundedness only by verificationism, "a semantic thesis that has the convenient property of entailing that psychologically inaccessible beliefs are ipso facto nontruth-valuable [sic]" (p. 124).

Without some form of verificationist semantics or other retreats from realism, Fodor points out, significant conceptual constraints will imply epistemic boundedness. "All cognitive psychologies thus far proposed," Fodor declares, impose such constraints (p. 125). Nor is this likely to change. "The point," according to Fodor,

is that any psychology must attribute some endogenous structure to the mind (really unstructured objects -- bricks, say -- don't have beliefs and desires and they don't learn things). And it's hard to see how, in the course of making such attributions of endogenous structure, the theory could fail to imply some constraints on the class of beliefs that the mind can entertain. (ibid.)

Part of the implication seems plausible enough. Fodor's claim a little later on that "as long as the class of accessible concepts is endogenously constrained, there will be thoughts that we are unequipped to think," is surely defensible, at any rate. (ibid.) But why must the attribution of an endogenous structure to the mind constrain the concepts available to it? Aside from a footnote to Aquinas' version of the Aristotelian argument (as paraphrased by Geach), Fodor gives us no further explanation. Perhaps it would help to look at the original.

III

The Aristotelian account seems unpromising at first. It depends on the communication of forms theory of sensory perception. On this theory, perception involves the transmission of a sensible form from the object to the sense organ (usually) via a medium. When George sees the red ball, both the ball and George's eye share a form; both the ball and the eye are literally, occurrently red. On a Lockean account, by contrast, George's sensation is the end of a causal chain beginning with the ball, and it is the only component of the chain that is occurrently red.

It is crucial for the Aristotelian account that the sense organ be potentially whatever quality it is to be sensitive to (De Anima 418a3). This requires that, prior to sensing, the organ actually possess none of those sensible qualities it is supposed to respond to. The eye must be in itself uncolored, the tongue neither moist nor dry, the flesh² neither hot nor cold, and so on.

Understanding is like perceiving for Aristotle, in that the communication of forms explains it as well. So, he infers, "as that which is capable of perceiving is to the objects of perception, so must be the intellect similarly to its objects" (429a13 - a17). Here Anaxagoras' point applies: the intellect, which can think all things, must be potentially all things (cf. 431b21), and so actually none of them. The intellect must potentially have all forms, and actually have none. The intellect, Aristotle proclaims, "must have no other nature than this, that it is potential" (429a20).³

Aristotle concludes that the intellect cannot be a material thing, a composite of matter and form, not because it cannot be material per se, but

because it can have no form. Thus the argument would tell equally against a dualism which differed from the view that the intellect is a material substance only in having it made of different stuff. Aristotle turns next to other related arguments that the intellect cannot be material, and there is a tendency in the medieval Aristotelians to focus on this aspect of Anaxagoras' insistence that the mind be "unmixed"; but it is clear that if Aristotle's argument shows anything, it shows that the intellect is no sort of substance, or at least no sort of intelligible substance, either material or immaterial.

So far, Fodor's argument is simply Aristotle's turned on its head. Fodor's claim that no psychology can escape epistemic boundedness, since any psychology must impose an endogenous structure on the mind, is not limited to materialist psychologies or structures, either. Also the possibility that the mind may possess an unintelligible form recalls Fodor's claim that, to the extent that cognition is not modular, it is probably inexplicable (Fodor, pp. 126 -129). There seems no great leap in equating 'forms' and 'structures', so that both Aristotle and Fodor wind up accepting the claim that if the intellect has an endogenous structure, then it is incapable of certain thoughts. They differ only in that Aristotle's modus tollens has become Fodor's modus ponens.

The difficulty in attributing Aristotle's argument for the Anaxagorean point to Fodor, of course, is that it is based on an archaic theory of perception, involving as it does the communication of forms, which neither Fodor nor any psychologist today would accept. Or would they?

On the standard reading, communication of forms theories of perception were displaced by representative theories sometime in the seventeenth century. The crucial difference was that representative theories do not require, as do communication theories, that the objects of sensation share the intrinsic

qualities of their sensations. Instead, the objects in themselves are supposed to be bereft of sensible forms such as colors, sounds, odors, tastes, and feels. Our perceptions represent their objects by modeling them through their internal relations, not by qualitatively resembling the objects in their sensory contents. Perceptions capture the structural relations of their objects, not their intrinsic character.

Another way of putting all this, however, would be to say that it is the form and not the content that is communicated in perception. This would be misleadingly anachronistic, admittedly, since sensible forms just were the forms of sensory contents such as red. But it would be neither misleading nor anachronistic, I would argue, to view modern representative theories as differing from their predecessors in their reduction of qualitative to quantitative forms, and not in a rejection of the communication of forms. However this works out as history, there seems nothing in the claim that perception or understanding involves a transmission of forms, where 'form' is used in its current sense as a synonym for 'structure', to which Fodor or most psychologists could object.

Unfortunately, this version of the communication theory won't work in Aristotle's argument for the formlessness of the intellect. If the intellect's taking on the form of a tiger means only that it represents one, and not that it becomes one (albeit intentionally), then the formlessness of the understanding demanded by the communication theory means only that the intellect in itself represents nothing, and not that in itself it is nothing, i.e., that it is not a substance. It might even be a material substance, provided that a material substance can be a universal representer. Digital computers, as instantiations of Turing machines, seem to be such substances.

The point can be made by reference to Aristotle's famous comparison of the potential intellect to a blank tablet which potentially contains all inscriptions (429b29). Aristotle understood the analogy quite differently from Locke. Aristotle introduces the analogy only to explain what he means by the potential of the potential intellect. For Aristotle, the activation of this potential involves more than the inscription of the name of something on the blackboard of the intellect: the intellect must become another instance of the species understood, in the sense that it actually takes on its form. The possession of any form at all by the intellect would mean that it does not have the potential to take on that form, or forms incompatible with it, and so would limit the scope of understanding.

But on the modern view the potential intellect is just like the blank slate, a representational medium. The communication of forms now means the modification of the structure of the intellect to match that of the object of cognition. On Fodor's view this would involve tokenings in the language of thought. My point is that the ascription of an endogenous structure to a representational medium need not limit what it can represent. There of course will be what Fodor calls "boring parametric considerations" (p. 121): the size of the slate places an upper bound on the length of possible inscriptions, for example. But there is no reason to expect these parametric limitations to translate into categorical limitations in the content of what can be represented. For instance, the materiality of the mind need pose no barrier to its taking on either material or immaterial forms, if this means only that it adopts a certain representational structure, rather than that it actually becomes an immaterial thing.

The upshot then is that the Aristotelian argument for Fodor's claim that the ascription of any endogenous structure to the intellect would necessarily circumscribe the scope of its understanding does not work under a representative theory of mind. There may be other a priori arguments against the possibility of a universal representative medium, perhaps involving Godelian self-reference.⁴ We might also inquire whether specific theses on the structure of minds -- materialism, computationalism, and the like -- entail conceptual constraints.

IV

In Elbow Room, Dennett proposes an argument for human sphexishness based on a computational view of mind that proceeds through something like epistemic boundedness. It goes like this (Dennett 1984, pp. 28 - 29):

1. Avoiding sphexishness requires responding to semantic features.
2. Physical systems can respond only to syntactical features.
3. Semantics is not determined by syntax.
4. Human minds are physical systems.
5. Therefore, human minds cannot avoid sphexishness.

None of the premises is uncontroversial, naturally, but I doubt whether there is a single understanding of 'syntax' and 'semantics' throughout the first three premises.

In the first premise, 'semantic' is used in the same broad sense as when Dennett says that "brains are meaning manipulators, information processors, or, as I shall say, semantic engines" (Dennett 1984, p. 28). Here it has no intrinsic connection to language, but could perhaps be replaced by 'evidentiary' or 'inferential'. The second premise is stated by Dennett thus: "physical mechanisms [are] syntactic engines, responding only to structural or formal

properties" (ibid.). Clearly again the sense of 'syntactic' is not connected to language, but how it is to be understood is unclear to me. The third premise, however, although a truism when both 'syntax' and 'semantics' are understood in the usual way as aspects of language, is not obvious when the terms are used in the extended senses of the first two premises. One reason that syntax doesn't determine semantics, for example, is that context is important. But this distinction uses 'syntax' in a narrower sense than would be plausible in the second premise. If 'syntactic' means just physical, and 'semantic' evidentiary, on the other hand, it is not obvious that syntax doesn't determine semantics. Isn't the fact that 'smoke' means fire determined by physical features?

One reason a syntactic emulator would lag behind a semantic engine is quantitative. A finitely axiomatizable grammar will have only countably many expressions. But there are uncountably many truths, not to say meanings. (Consider real number theory.) So if my mind is a syntactic engine in this sense -- a finitely axiomatizable automatic formal system⁵ -- then there will be uncountably many meanings it cannot capture, meanings which presumably would be directly apprehended by a semantic engine. But this is just epistemic boundedness, quod erat demonstrandum.

Not quite. A condition of epistemic boundedness was that the unthinkable thoughts be "epistemically significant" (Fodor, p. 120). Only if the inaccessible meanings were significant would their inaccessibility imply sphexishness, too. So both the epistemic boundedness and the sphexishness claims await some reason to think that the inaccessible meanings are not so due to their outrageous complexity or other "boring parametric considerations." So this is just another cardinality argument, which assumes that perfect

intelligence must be infinite, when surely finite intelligence is all finite creatures require.

* * * * *

There are undoubtedly an infinity of truths which lack expressions in Mentalese because they are too long or complex. Epistemic boundedness and cognitive sphexishness seem to be claiming more than this, however. It may well be that the endogenous structure of our minds makes us all Mentalese aphasics as dramatic, say, as the man who mistook his wife for his hat, and that only the universality of the phenomenon and the tendency of the mind to paper over its blind spots keeps the dread secret hidden. There seems little reason to think so a priori, however. At least it is hard for me to see how it is hard for Fodor to see how it could be otherwise.⁶

NOTES:

1. Fodor doesn't realize how close he is to Aquinas, because he reverses the direction of the Angelic Doctor's reasoning: he takes him to be arguing from the immateriality of the intellect to its universality, instead of the converse.
2. Or whatever is the proper organ of touch: see 422b34ff.
3. This actually describes only what came to be known as the potential intellect; in the following chapter, Aristotle notoriously proposes as well an intellect whose essence is pure activity.
4. Cf. Hofstadter pp. 534 - 36 on the relevance of the halting problem to the sphexishness issue.
5. Cf. Haugeland 1986 Chapter 2.
6. Read to the North Carolina Philosophical Society and to the Faculty Research Group at Davidson College. Thanks to Christopher Bell, Irwin Goldstein, Richard E. Grandy, Michael Resnik, and Richard Smyth for helpful comments.

BIBLIOGRAPHY

- Cottingham, John et al. The Philosophical Writings of Descartes Vol I. Cambridge University Press, 1985.
- Dennett, Daniel. Brainstorms, MIT Press, 1978.
- Dennett, Daniel. Elbow Room. MIT Press, 1984.
- Fodor, Jerry. The Modularity of Mind. MIT Press, 1983.
- Hamlyn, D.W. Aristotle's De Anima, Books II, III. Oxford University Press, 1968.
- Haugeland, John. Artificial Intelligence: The Very Idea. MIT Press, 1986.
- Hofstadter, Douglas. Metamagical Themas. Basic Books, 1985.
- Wooldrige, Dean. The Machinery of the Brain. McGraw Hill, 1963.