Between Saying and Doing: Towards an Analytic Pragmatism

Lecture Three: Artificial Intelligence and Analytic Pragmatism

Section 1: AI-Functionalism

Classical AI-functionalism is the claim that there is some program such that anything that runs that program can pass the Turing test, that is, can deploy a vocabulary in the same sense in which any other language-users do. And that is to say that a computer language, in which any such algorithm can be expressed, is in principle VP-sufficient to specify abilities that are PV-sufficient to deploy an autonomous vocabulary. So in my terms, classical AI-functionalism claims that computer languages are in principle sufficient *pragmatic metavocabularies* for some autonomous vocabulary. Since computer languages (syntactically, context-free languages) are not themselves autonomous vocabularies, the basic claim of AI-functionalism is a *strict expressive bootstrapping* claim about computer languages as pragmatic metavocabularies for much more expressively powerful vocabularies: natural languages. It is a claim about what one must be able to *do* in order to count thereby as *saying* anything.

I take the working-out of various forms of *functionalism* in the philosophy of mind (including the computational species)—a view about what is expressed by (that is, about the semantics of) intentional vocabulary—to be one of the cardinal achievements of Anglophone philosophy in the second half of the twentieth century. (Behaviorism was a larval stage of it.) It deserves to be thought of as a third core program of the classical project of philosophical analysis, alongside empiricism and naturalism.

Section 2: Classic Symbolic Artificial Intelligence

Functionalism is a more promising explanatory strategy when addressed to *sapience* rather than to *sentience*.

The slogan of classical symbolic AI is: *Mind is to brain as software is to hardware*.

Section 3: A Pragmatic Conception of Artificial Intelligence

What I will call the "algorithmic pragmatic elaboration" version of AI-functionalism—or just "pragmatic AI"—is the claim that there is a set of practices-or-abilities meeting two conditions:

1) It can be algorithmically elaborated into (the ability to engage in) an autonomous discursive practice (ADP);
and
2) Every element in that set of primitive practices-or-abilities can intelligibly be understood to be engaged in, possessed, exercised, or exhibited by something that does not engage in any ADP.

The first of these is a kind of PP-sufficiency claim—specifically, an algorithmic elaboration PP-sufficiency claim. The second is the denial of a PP-necessity claim.

The important difference between this formulation and classical symbolic AI is that the connection to computers (or as I would prefer to say, automata) is not established via the principle that computers are symbol-manipulating engines and that according to the computational theory of the mind thinking just consists in manipulating symbols, but rather via PP-sufficiency of the algorithmic elaboration sort that I discussed in my second lecture.

The pragmatic version of AI is a very general *structural* claim that could be made about any kind of practice-or-ability. The issue it raises need have nothing whatever to do with symbol-manipulation. For *any* practice-or-ability *P*, we can ask whether that practice-or-ability can be algorithmically *decomposed* (pragmatically analyzed) into a set of primitive practices-or-abilties such that:

1) They are PP-sufficient for *P*, in the sense that *P* can be algorithmically elaborated from them (that is, that *all* you need in principle to be able to engage in or exercise *P* is to be able to exercise those abilities plus the algorithmic elaborative abilities, when these are all integrated as specified by some algorithm);
and
2) One could have the capacity to engage in or exercise *each* of those primitive practices-or-abilities without having the capacity to engage in or exercise *P*.

If those two conditions are met, we may say that *P* is *substantively algorithmically decomposable* into those primitive practices-or-abilities.

Section 4: Arguments Against AI-Functionalism: Ranges of Counterfactual Robustness for Complex Relational Predicates

Standard arguments against classical symbolic AI diagnose it as built around the traditional platonist or intellectualist commitment to finding some bit of explicit knowing- (or believing-) *that* behind every bit of implicit practical knowing-*how*. By contrast, the corresponding argument against the substantive practical algorithmic decomposability version of AI would have to offer reasons for pessimism about the possibility of algorithmically resolving essentially discursive knowing- (or believing-) *that* without remainder into non-discursive forms of knowing-*how*. Whatever problems there may be with this kind of AI, they do not stem from some hidden *intellectualism*, but concern rather the particular variety of *pragmatism* it articulates: *algorithmic pragmatism* about the discursive. For what makes the substantive algorithmic practical elaboration model of AI interesting is the relatively precise shape that it gives to the pragmatist program of explaining knowing-that in terms of knowing-how: specifying in a non-intentional, non-semantic vocabulary what it is one must *do* in order to count as deploying some vocabulary, hence as making intentional and semantic vocabulary applicable to the performances one produces (a kind of expressive pragmatic bootstrapping).

To argue against the practical algorithmic elaboration version of AI we must find some aspect exhibited by all autonomous discursive practices that is not algorithmically decomposable into non-discursive practices-or-abilities. That would be something that is PV-necessary for deploying any autonomous vocabulary (or equivalently, PP-necessary for any ADP) that cannot be algorithmically decomposed into practices for which no ADP is PP-necessary.

The productivity of language guarantees that anything that can talk can form predicates specifying an indefinitely large class of relational properties. The problem is that doxastic updating for language-users requires distinguishing among all of these, those that are from those that are not relevant to the claims and inferences one endorses—that is, those which fall within the range of counterfactual robustness of those claims and inferences. And it is not plausible that *this* ability can be algorithmically decomposed into abilities exhibitable by non-linguistic creatures.

The argument:

- One cannot talk unless one can ignore a vast variety of considerations one is capable of attending to, in particular those that involve complex relational properties, that lie within the range of counterfactual robustness of an inference.
- Only something that can talk can do that, since one cannot ignore what one cannot attend to, and for many complex relational properties, only those with access to the combinatorial productive resources of a language can pick them out. No non-linguistic creature can be concerned with fridgeons or old-Provo eye colors.
- So language-use, deploying autonomous vocabularies, brings with it the need for a new kind of capacity: for each inference one entertains, to distinguish in practice among all the new complex relational properties that one comes to be able to consider, those that are, from those that are not relevant to assessing it.
- Since non-linguistic creatures have no semantic, cognitive, or practical access at all to most of the complex relational properties they would have to distinguish to assess the goodness of many material inferences, there is no reason at all to expect that that sophisticated ability to distinguish ranges of counterfactual robustness involving them could be algorithmically elaborated from the sort of abilities those creatures do have.

Section 5: Practical Elaboration by Training

We do not need to assume that discursive practice is substantively algorithmically decomposable into non-discursive practices-or-abilities, on pain of making entering into those practices and acquiring those abilities—by us as a species, and as individuals—unintelligible, because there is another sort of PP-sufficiency relation besides algorithmic elaboration: practical elaboration by training. This occurs when, as a matter of contingent empirical fact, there is a course of practical experience or training that will bring those who have one set of abilities to have another set of abilities. We need to acknowledge this sort of PP-sufficiency in any case, in order to account for the provenance of the abilities treated as primitive for the purposes of algorithmic elaboration. Wittgenstein urges us to see them not only as crucial for the advent of discursive practices-or-abilities, but also as pervasive within up-and-running discursive practices, alongside algorithmic elaboration.

Section 6: Algorithmic Pedagogical Decomposition and Pedagogical Politics

For the Wittgensteinian pragmatist, appeal to algorithmically nondecomposable, contingent, parochial abilities is compatible with investigating PP-sufficiency and PP-necessity *dependency*

relations between such abilities and practices, as well as the PV- and VP-sufficiency relations they stand in to vocabularies. One analytic issue raised by the consideration of what I will call *pedagogic* practical elaboration and decomposition is a broadly *political* one.

To have a *solved pedagogical problem* with respect to an output practice-or-ability is to have an *empirically sufficient conditional branched training regimen* for it. This is something that as a matter of contingent fact can take any novice who has mastered the relevant range of primitive practical capacities, and by an algorithmically specifiable Test-Operate-Test-Exit cycle of responses to their responses *as a matter of fact* get that novice to catch on to the target ability.

There are two basic attitudes (defining a spectrum between them) that one might have toward any target ability for which we do not have a pedagogical algorithm codifying a complete solution to the training problem:
Empirical-hierarchical: It is just a brute empirical fact that people not only have different abilities, but are in important respects more or less able. With respect to any sort of target ability, some are more trainable, better learners, than others. The training regimen not only inculcates or elicits the skill that is its target, but along the way sorts candidates into those who can, and those who cannot learn or be trained in it, as well as into those who learn it faster or more easily—measured by how long or how many steps it takes to get them through the pedagogical flowchart to the exit of that practical labyrinth. On this view, it is compatible with just dealing, and perhaps even constitutive of a dimension of justice, for an institution to factor this sort of second-order ability into its reward-structure.
Utopian-egalitarian: The hierarchical sorting of candidates into more or less trainable is relative to the training regimens that happen to be available. Different regimens might produce quite different rankings. If that is so, and the fact that we have actually implemented one set of training procedures rather than another is quite contingent, conditioned by adventitious, in-principle parochial features of the actual history of the training institutions, then the inferences from the actual outcomes of training either to the attribution of some kind of *general* second-order ability or to the justice of rewarding the *particular* sort of second-order ability that really *is* evidenced thereby—merely being more (or more easily) trainable by the methods we happen to have in place—are undercut. Our failure to provide a more comprehensive set of training alternatives, to have filled in the pedagogic flow-chart more fully, ultimately, to have completely solved the relevant training problem, should be held responsible for the training outcome, rather than supposing that sub-optimal outcomes reveal evaluatively significant deficiencies on the part of the trainee. At the limit, this attitude consists in a *cognitive* commitment to the effect that there is in principle a complete pedagogic algorithmic solution for every target skill or ability to the possession of which it is just to apportion rewards, and in a *practical* commitment to find and implement those solutions.