May 14, 2006

Between Saying and Doing: Towards an Analytic Pragmatism

Lecture 3 (May 17, 2006):

Artificial Intelligence and Analytic Pragmatism

Section 1: AI-Functionalism

The thought with which I introduced meaning-use analysis, and the paradigm of a pragmatically mediated semantic relation, arises when we put together two sorts of story:

- An account of what one must *do* in order to count as *saying* something—that is, of some practices-or-abilities that are PV-sufficient to deploy a vocabulary,

  and

- A characterization of another vocabulary that one can use to *say* what it is one must *do* to be doing something, for instance, in order to be *saying* something else—that is, of a vocabulary that is VP-sufficient to *specify* the practices-or-abilities, which may be PV-sufficient to deploy another vocabulary.

When we compose these, the resultant meaning-use relation is the relationship between vocabularies that I have called the "pragmatic metavocabulary" relation. I have suggested that this relation is most illuminating when the pragmatic metavocabulary is demonstrably *expressively weaker* than the vocabulary for which it is a pragmatic metavocabulary. This is what I have called "pragmatic expressive bootstrapping," in the strict sense. We have seen several examples of this phenomenon:

- Syntactic pragmatic bootstrapping, within the Chomsky hierarchy of grammars and automata, in that *context-free* vocabularies are VP-sufficient to specify Turing machines (two-stack pushdown automata), which are in turn PV-sufficient to deploy (produce and recognize) *recursively enumerable* vocabularies.
- I argued (proleptically in my first lecture, and then in further detail in the Appendix to the second lecture) that *non*-indexical vocabulary is VP-sufficient to specify practices PV-sufficient to deploy *indexical* vocabulary.
- I have mentioned, though not discussed, Huw Price's pragmatic naturalism—my term, he talks about it as "subject naturalism", by contrast to "object naturalism"—which denies semantic reducibility or even supervenience, but seeks to lessen the sting of that denial by specifying in a naturalistic vocabulary what one must *do* in order to deploy various irreducibly non-naturalistic vocabularies—e.g. normative or intentional ones.

I will argue in later lectures that *normative* vocabulary is a sufficient pragmatic metavocabulary for *modal* vocabulary: a case where the expressive ranges are at least impressively *different* even if not rankable as expressively weaker and stronger.

In this lecture, I will discuss another philosophically significant contention of this sort: the claim, thesis, or program that is usually associated with the rubric "artificial intelligence." (As workers in the field fully understand, the phrase "artificial intelligence" is a terrible way to pick out the topic. Artificial intelligence is to be real intelligence, created by artifice. But artificial diamonds are not real diamonds created by artifice. They are fake diamonds. Real diamonds created in a laboratory are *synthetic* diamonds. And what is at issue is not *intelligence*—a phenomenon that admits of degrees and has its primary application to comparative assessments *within* the discursive community. It is really *sapience* that is at issue—something we language-users have and cats do not. So the issue would be better identified as "synthetic sapience" than "artificial

intelligence." But it is too late to get the label right.) Very crudely, AI is the claim that a computer could in principle *do* what is needed to deploy an autonomous vocabulary, that is, in this strong sense, to *say* something. It is accordingly a thesis about meaning-use relations, in my sense. The classical Turing test for the sort of 'intelligence' at issue is a *talking* test; something passes it if one cannot tell it from a human speaker—that is, from someone who engages in autonomous discursive practices, someone who deploys an autonomous vocabulary—by talking to it. 'Intelligence' in this sense just consists in deploying such a vocabulary. 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. Now I take it that computer languages are not themselves autonomous vocabularies. For such context-free languages lack essential kinds of vocabulary. We cannot make sense of linguistic communities that speak only Fortran or C^{++} (though some groups of engineers, when talking among themselves, on occasion come close). Insofar as that is right, the basic claim of AI-functionalism is an *expressive bootstrapping* claim about computer languages as pragmatic metavocabularies for much more expressively powerful vocabularies: natural languages. It deserves a prominent place on the list of philosophically significant pragmatic expressive bootstrapping claims I just offered. And it should be a principal topic of philosophical meaning-use analysis. So let us see what the meaning-use-analysis metavocabulary I have been deploying can help us understand about it—what lessons applying these metaconceptual tools can teach us.

Of course, AI has not traditionally been thought of as an expressive bootstrapping claim about a pragmatic metavocabulary. How could it have been? But although its twentieth-century version developed later than the others, functionalism in the philosophy of mind, including its central computational species, deserves to be thought of as a third core program of the classical project of philosophical analysis, alongside empiricism and naturalism. (For reasons I indicated last time, I think of behaviorism as the larval stage of functionalism. Looking back from the vantage-point achieved by the development of functionalism, we can see that every advantage or consideration speaking in favor of behaviorism also does so for functionalism. And making the move from pragmatic elaboration by single-state transducing automata to multistate ones adds a great deal of theoretical flexibility and explanatory power. The *only* reasons to hold out for behaviorism would be instrumentalist qualms about the positing of theoretical entities in the form of functional states.) And since, as I have indicated, that is, about the relation between practices or abilities and the deployment of vocabularies, insofar as functionalist successors to behaviorist programs in the philosophy of mind do deserve a prominent place at the analytic table, that fact indicates that the sort of broadening of the analytic semantic project to include pragmatics that I have been recommending has in fact implicitly been under way for some time.

Section 2: Classic Symbolic Artificial Intelligence

I take the working-out of various forms of functionalism in the philosophy of mind to have been one of the cardinal achievements of Anglophone philosophy in the second half of the twentieth century. Neither it nor its sophisticated species, AI-functionalism, is the product of a single thinker.

Even more than the other core programs of twentieth century empiricism and naturalism, functionalism has grown up gradually, with different parts of the picture being filled-in by different thinkers, in different disciplines—computer scientists, psychologists, and other cognitive scientists at least as much as philosophers. One of the things I think we have found out along the way is that functionalism is a more promising explanatory strategy when addressed to *sapience* rather than to *sentience*—when it is addressed to our understanding of states such as belief, rather than pains or sensations of red (themselves importantly different cases, as the different grammar of the two phrases indicates). In broadest terms, the basic idea of functionalism is to assimilate bits of intentional vocabulary such as "belief that *p*" to terms classifying something in terms of the role it plays in a more complex system. So the relations between 'belief', 'desire', 'intention', and 'action' might be modeled on the relations between 'valve', 'fluid', 'pump', and 'filter'. The most immediate attraction of such an approach is the *via media* it provides between the traditional alternatives of materialism and dualism. All valves, that is, all things playing the functional role of a valve in any system, are physical objects, and they can function as valves only in virtue of their physical properties. So far, *materialism* was right: functional vocabulary applies exclusively to physical objects. But what valves have in common that makes that term properly apply to them is not a physical property. Mechanical hydraulic valves, heart valves, and electronic valves may have no physical properties in common that they do not share with a host of non-valves. So far, *dualism* was right: functional properties are not physical properties. *Automaton* functionalism is a species of this general view that looks specifically at the functional roles items can play in multistate transducing automata. By the term 'AI-functionalism' I shall mean automaton functionalism about *sapience*—about what it is in virtue of which *intentional-state* vocabulary such as "believes that" is applicable to something, that is, in the terms I have been using (and which are endorsed by appeals to the Turing test), the capacity to engage in any autonomous discursive

practice, to deploy any autonomous vocabulary, to engage in any discursive practice one could engage in though one engaged in no other. (This is the view we have in mind when we think of the transition from behaviorism to functionalism in the philosophy of mind as corresponding to the transition from modeling the mind on single-state stimulus-response transducing automata to modeling it on multistate transducing automata.)

So understood, AI-functionalism admits of different interpretations. Approaching it as asserting a particular kind of pragmatically mediated semantic relation between vocabularies—as making an expressive bootstrapping claim about a particular kind of pragmatic metavocabulary for some autonomous vocabulary—as meaning-use analysis suggests, leads to a characterization that is in important ways broader than traditional formulations. So I want to say something about that difference.

I will call what I take to be the received understanding of the central claims of AI, what John Searle calls the "strong thesis of AI"[1], "classical symbolic AI"—or just "classy AI", for short. Here is how I understand it.[2] Its slogan is: "*Mind is to brain as software is to hardware*." It sees a crucial difference between modeling the mind on computer programs and all previous fashionable, rashly enthusiastic claims that some bit of impressively powerful new technology would also, *inter alia*, give us the key to unlock the secrets of the mind—telephone switchboards, clockworks, and if we go far enough back, even potter's wheels having been taken to play that role. For computing is manipulating symbols according to definite rules (the

[1] Searle "Minds, Brains, and Programs" [ref.]

[2] I think the characterization that follows would be very widely endorsed as expressing the core motivations of the view at issue, both by proponents and opponents of classy AI. My presentation is guided by that of my former colleague, John Haugeland, in his excellent introduction: *Artificial Intelligence: The Very Idea* [Bradford Press, MIT, [ref.]].

algorithms implicit in automaton state-tables). And, the claim is, thinking or reasoning, the fundamental sort of operation or activity that constitutes sapience, just *is* manipulating symbols according to definite rules. This *computational theory of the mind* is the basis of the argument for AI-functionalism. It is a view that long antedates the advent of computers, having been epitomized already by Hobbes in his claim that "reasoning is but reckoning."[3]

Now the plausibility of understanding thinking as symbol-manipulation at all depends on taking symbols to be more than just sign-designs with a certain syntax. They must be *meaningful*, semantically contentful signs, whose proper manipulation—what it is *correct* to do with them—must depend on the meanings they express, or on what they represent. Traditionally, this fact meant that there was a problem reconciling the computational view of the mind with naturalism. Physics does not find meanings or semantic properties in its catalogue of the furniture of the world. They are not, or at any rate not evidently, physical properties. So how could any physical system *be* a computer, and respond differentially to signs depending on the *meanings* they express? Looking back from the vantage point vouchsafed us by the development of actual computing machinery—and the realization that doing numerical calculation by the algorithmic manipulation of numerals was only one instance of a more general symbol-manipulating capacity—provided a possible answer. Already for Descartes, the thorough-going isomorphism he had established between algebraic formulae and geometric figures suggested that manipulating the formulae according to the rules proper to them could, not just express, but *constitute* or *embody* an understanding of the figures. The isomorphism amounts to an encoding of semantic properties in syntactic ones. A physical system can accordingly be a computer—manipulate symbols in ways that accord with their meanings—

[3] [ref.—Get it from Ch. 1 of *AIVI*.]

because such an encoding ensures that, in Haugeland's slogan, if the automaton takes care of the syntax, the semantics will take care of itself.

Usually, what you get when you manipulate symbols in ways that exploit isomorphisms to what they are symbols of is a *simulation*. Computers can manipulate symbols so as to model traffic patterns, weather systems, and forest fires. No one is liable to confuse the symbol-manipulating with the phenomena it simulates—the computation with the traffic, the weather, or the fire. But AI-functionalism claims that unlike these cases, manipulating symbols in ways that suitably respect, reflect, and exploit isomorphisms with what those symbols for that very reason count as expressing or representing is not just a *simulation* of thinking, but *is* thinking itself. That is what it *is* to deploy a vocabulary *as* a vocabulary, that is, as meaningful. The *only* reason for according thought this uniquely privileged position—as the one phenomenon that cannot be symbolically *simulated* without thereby being actually *instantiated*—is whatever reason there is to think that the symbolic-computational theory of the mind is correct. And that is a very substantive, potentially controversial theory, with a correspondingly large burden of proof.

Section 3: A Pragmatic Conception of Artificial Intelligence

I think that symbolic AI's focus on the Turing test is appropriate. There is just no point in insisting that something that is genuinely indistinguishable from other discursive practitioners in conversation—no matter how extended and wide-ranging in topic—should nonetheless not be counted as *really* talking, so thinking (out loud), and deploying a meaningful vocabulary. But although the slide can seem unavoidable, it is a long way from acknowledging the criterial

character of the Turing test to endorsing the computational theory of the mind on which classical symbolic AI is predicated. The line of thought I have just rehearsed invites a focus on the issue of the *symbolic* character of thought that I think is misleading. And for that reason it mislocates, as it seems to me, what really is the most important issue in the vicinity: the claimed *algorithmic* character (or characterizability) of thought or discursive practice.

In my second lecture, I argued that from the point of view of meaning-use analysis, the principal significance of automata does not lie in their capacity to manipulate symbols, but rather in their implementing a distinctive kind of PP-sufficiency relation. Multistate transducing automata *algorithmically elaborate* a set of primitive abilities into further abilities—abilities which, just because they can be so exhibited, can then be regarded as complex, as pragmatically *analyzable* into those primitive abilities plus the basic algorithmic elaborating abilities. This characterization of automata suggests that AI be understood broadly as a claim to the effect that such an analysis or decomposition is possible of some autonomous discursive practice—the practice-or-ability to deploy some vocabulary that can be deployed though one deploys no other. That is, it claims that some autonomous discursive practice can be exhibited as the algorithmic elaboration of a set of primitive abilities, which are accordingly PP-sufficient for that ADP.

That claim by itself would not be interesting or controversial. For the null elaboration is also an algorithmic elaboration (albeit a degenerate one). So the condition would be trivially satisfied, just because there *are* autonomous discursive practices-or-abilities. What is needed to turn the claim that some set of primitive abilities can be algorithmically elaborated so as to be PP-sufficient for some autonomous discursive practice into a substantial one is a further

constraint on the primitive abilities. Given the reasons for being interested in AI-functionalism in the first place, what we want is to stipulate that what are to be counted as primitive abilities with respect to such an algorithmic elaboration must themselves in some sense not already be *discursive* abilities.

Here is the version that I propose. What I will call the "algorithmic pragmatic elaboration" (APE) 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.[4]

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.

This approach to AI-functionalism shifts the focus of attention away from the role of *symbols* in thought, away from the question of whether thinking just *is* manipulation of symbols, and away from the issue of whether isomorphism is sufficient to establish genuine ('original', rather than merely 'derivative') semantic contentfulness. It is true that I am here still thinking of what is at issue in sapience as a matter of deploying *vocabularies*, that is, using symbols, semantically

[4] Note that this condition does *not* require that one be able to engage in or exercise each of these primitive practices-or-abilities *individually*. Perhaps they come in clusters that mutually presuppose one another. The condition will still be met so long as each such cluster does not presuppose the capacity to engage in any ADP.

significant signs—not in a derivative way, but in whatever way is fundamental in the sense of being exhibited by *autonomous* discursive practices-or-abilities, and the vocabularies they deploy. But—and here is the important difference from classical symbolic AI—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 previous lecture. And the structural question AI-functionalism asks is an issue that can arise for *any* ability—not just those that involve symbol-use. That is, 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 engage in 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. So, for instance, the capacity to do long division *is* substantively algorithmically decomposable, into the primitive (with respect to this decomposition) capacities to do multiplication and subtraction. For one can learn to multiply, or again, to subtract, without yet having learned how to divide. Perhaps the capacity to play the

piano is like this, since one can learn how to finger each key individually, and to adjust the intervals between doing so. By contrast, the capacities to respond differentially to red things and to wiggle my index finger probably are not substantively algorithmically decomposable into more basic capacities. These are not things that I do *by* doing something else. If I do not have those abilities, there is no way to put them together as the complex results of some structured sequence of other things—even with the flexibility of conditional branched schedule algorithms, hence of Test-Operate-Test-Exit feedback loops of perception, action, and further perception of the results of the action. The abilities to ride a bicycle, to swim, or to hang-glide might or might not be substantively practically algorithmically decomposable, and the empirical question of whether or not they are, and if so, how, is of considerable pedagogical significance (about which more later).

In sorting these examples, I have marked the classifications as tentative, by using qualifiers such as 'probably'. For it is important to realize that the question of which practices-or-abilities are and which are not substantively algorithmically decomposable is an *empirical* question, not the sort we can confidently answer from our armchairs. It may be, for instance, as Hornsby has pointed out, that the only way I can contract some muscle in my left hand to a particular degree is by playing a certain note on the violin. And even in the case of responding differentially to visible red things or wiggling my finger, we will be able to describe *some* algorithmic practical decompositions. For instance, the capacity to distinguish red things can be algorithmically assembled out of or elaborated from the ability to sort objects into similarity classes based on (as we would presystematically put it) shape-and-color (distinguishing red cubes from green spheres), color-and-size (distinguishing large red things from small green ones) and shape-and-size (distinguishing large spheres from small cubes). But in that case one suspects—but I would not claim to know—that the practical algorithmic decomposition would not be *substantive*. That is, it would probably turn out that the only creatures who could exhibit those abilities also already have the ability to sort things just by color. Hornsby's case, by contrast, is one where the corresponding fact plausibly does *not* hold. What is at issue is contingent, parochial PP-necessity relations, and actual empirical investigation is

required to determine what abilities are and are not dependent, for creatures like us, on others. For that turns on how we actually work, on how these practical recognitive and responsive abilities are implemented in us—not just at a personal level, but also at sub-personal levels. Thus one way of investigating the sort of PP-necessity relations involved in the requirement that a practical algorithmic decomposition be substantive is to look at organisms that have been damaged in various ways, for instance, by localized strokes or bullets in their brains. For in that way one can find out that a whole suite of abilities comes together—that as a matter of empirical fact, anyone who has lost the ability to do A will also have lost the ability to do B. What we can tell on *a priori* grounds—since the alternative is an infinite regress—is only that there must be some things that we can just do, without having to do them *by* doing something else, by algorithmically elaborating other abilities. (That is compatible with there being clusters of abilities such that we *can* do any of the component things *by* doing some—possibly conditioned—sequence of the others.)

So the question of whether some practice-or-ability admits of a substantive practical algorithmic decomposition is a matter of what contingent, parochial, matter-of-factual PP-sufficiencies and necessities actually are exhibited by the creatures producing the performances in question. That question is very general and abstract, but also both empirical and important. It is a very general *structural* question about the ability in question. That issue, as such, however, has nothing whatever to do with symbol-manipulation. My suggestion is that we think of the core issue of AI-functionalism as being of this form. The issue is whether *whatever* capacities constitute sapience, *whatever* practices-or-abilities it involves, admit of such a substantive practical algorithmic decomposition. If we think of sapience as consisting in the capacity to deploy a vocabulary, so as being what the Turing test is a test for, then since we are thinking of sapience as a kind of symbol use, the target practices-or-abilities will also involve symbols. But that is an entirely separate, in principle independent, commitment. That is why I say that classical symbolic AI-functionalism is merely one species of the broader genus of algorithmic

practical elaboration AI-functionalism, and that the central issues are mislocated if we focus on the symbolic nature of thought rather than the substantive practical algorithmic analyzability of whatever practices-or-abilities are sufficient for sapience.

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

Because (for reasons I have already indicated) the two stand or fall together, arguments against the plausibility of the claims of classic symbolic AI-functionalism take the form of arguments against the computational theory of the sapient mind. These arguments include doubts about the possibility of explicitly codifying in programmable, hence explicitly statable, *rules* all the implicit practical background skills necessary for thoughtful engagement with the world, challenges to the adequacy of the semantic epiphenomenalism inherent in treating syntactic isomorphism as sufficient for non-derivative contentfulness, and reminders of the sort epitomized by Searle's Chinese Room thought-experiment of how badly the essentially third-person point of view of this sort of functionalist successor to behaviorism fits with intuitions derived from our first-person experience of understanding, grasping meanings, and having contentful thoughts. Reasons for skepticism about the sort of AI understood instead as claiming the substantive algorithmic decomposability of autonomous discursive practices-or-abilities into non-discursive ones must take a distinctly different shape.

For instance, Dreyfus objects to classical symbolic AI on the grounds that it requires that all the implicit practical skills necessary for understanding our ordinary life-world have to be

made explicit in the form of rules (codified in programs). He diagnoses classy AI 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*. Like Dewey, he is skeptical about that framing commitment. 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, on the contrary, 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 pragmatic expressive bootstrapping).

What arguments are there against this pragmatist version of AI? One sort of complaint that might initially be tempting, but which seems to me not in the end to be well-taken, concerns the *form* in which practices-or-abilities are represented when we think about elaborating them by multistate transducing automata. The primitive abilities that serve as inputs to that process of practical algorithmic elaboration are of two kinds: discriminative abilities and performative abilities. That is, they are capacities to respond differentially to a determinate range of stimuli, and capacities to respond differentially by producing performances of some determinate kind. The idealizing assumption I called "response substitution" is that discriminative and performative abilities in the basic repertoire can be elaborated by being responsively linked in any arbitrary combination. The idealizing assumptions of arbitrary state formation and permutation are that those stimulus-response connections can

then be combined into automaton-states in any way at all. I have emphasized that transducing automata that elaborate such repertoires of basic abilities are not at all restricted to symbol manipulation. In the very broad sense appropriate to automata thought of as ways of elaborating practices-or-abilities, potential symbols are what can both be read and written, that is, responsively discriminated as stimuli and differentially elicited as responses. What can be symbols in this sense play no privileged role in practical algorithmic functionalism. From this point of view, the symbolic focus of classic AI seems *ad hoc* and unmotivated. But that conceptual broadening may seem still not to go far enough. The remnants of the behaviorism out of which practical algorithmic functionalism grows are still to be found in the stimulus-response character of the primitive abilities it shows how to elaborate. As I pointed out in the previous lecture, however, how substantial a restriction is it that a stimulus-response vocabulary is assumed to be VP-sufficient to specify all the primitive abilities that can serve as inputs to the process of algorithmic elaboration? Is that assumption by itself a reason to doubt the possibility of the practical algorithmic elaboration version of AI?

I do not think that it is. The insistence that a stimulus-response vocabulary is VP-sufficient does not represent a significant restriction on what can show up as primitive abilities as long as we do not put any *other* constraints on the ability-specifying vocabulary—that is, as long as we are allowed to help ourselves to the full expressive power of any autonomous vocabulary in specifying the discriminable stimulus-kinds and differentially elicitable response-kinds. The abilities in question are 'primitive' only with respect to the process of algorithmic elaboration. In every other sense they may be very sophisticated. The stimulus-response formulation by itself does not keep us from considering as 'primitive' capacities the abilities to keep ourselves at a suitable distance from a conversational partner, distinguish cubist paintings done by Braque from those done by Picasso, drive from New York to San Francisco, or build a house. The abilities considered as inputs to the process of practical algorithmic elaboration need not be at all mechanical. They could, for instance, be things that only *sentient* creatures could exhibit. What is at issue in practical algorithmic AI-functionalism is the substantive algorithmic decomposability of sapience, not at all whether sapience is possible without sentience (as some have thought was a core claim of classic symbolic AI). The *only* restriction on candidate primitive abilities that matters is the one imposed on them by the requirement that the algorithmic decomposition of an ADP be *substantive*. That is the requirement that each of the primitive abilities out of which the capacity to deploy an autonomous vocabulary is to be elaborated must be something that could be exhibited by a creature that does *not* deploy an autonomous vocabulary. That *is* a

substantial restriction—one that very well may, as a matter of empirical fact, rule out the abilities I just offered as examples of sophisticated discriminative and performative abilities that would otherwise qualify as 'primitive' for the purposes of practical algorithmic elaboration.

The form of the claim tells us that 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. I do not claim to have a knock-down argument here. But the best candidate I can think of to play that role is the practice of doxastic updating—of adjusting one's other beliefs in response to a change of belief, paradigmatically the addition of a new belief.

It is pretty clear that this set of practices-or-abilities is a PV-necessary aspect of the deployment of any vocabulary. For any set of practices to count as *discursive*, I claimed last time, it must accord some performances the significance of *claimings*. It is a necessary feature of that significance that what is expressed by those performances stands to other such contents in broadly *inferential* relations of being a reason for or against. That is, the practical significance of claiming includes undertaking a commitment that has other commitments and entitlements (or lack of entitlements) to commitments as its consequences, that can itself be a consequence of other commitments, and whose entitlement also depends on its relation to one's other commitments. One understands or grasps the content expressed by some bit of vocabulary that can be used to make claims only to the extent to which one can tell in practice (respond differentially according to) what follows from it and what it follows from, what other

commitments and entitlements the various commitments it can be used to undertake include and preclude. And that is to say that one understands what a bit of vocabulary means only insofar as one knows what difference undertaking a commitment by its use would make to what else the one using it is committed or entitled to—that is, insofar as one knows how to update a set of commitments and entitlements in the light of adding one that would be expressed using that vocabulary. Discursive understanding of this sort is a more-or-less affair. One need not be omniscient about the significance of a bit of vocabulary in order to deploy it meaningfully. But if one has *no* idea what practical consequences for other commitments a claim using it would have, then one associates no meaning with it at all.

If all that is right, then the question of whether doxastic updating can serve as a reason to be pessimistic about the practical algorithmic elaboration version of AI comes down to an assessment of the prospects for a substantive algorithmic decomposition of the ability to update. Why might one think that no such decomposition is possible—that is, that that essential discursive ability could not be algorithmically elaborated from any set of non-discursive abilities? The key point, I think, is that the updating process is highly sensitive to collateral commitments or beliefs. The significance of undertaking a new commitment (or relinquishing an old one) depends not just on the content of that commitment, but also on what else one is already committed to. I will argue in my next lecture that we can think of this global updating ability as a collection of sub-abilities: as the capacity, in one's actual doxastic context, to associate with each commitment a range of counterfactual robustness. To do that is to distinguish, for each commitment (including inferential commitments), which further commitments would, and which would not, infirm or defeat it. This includes not only claims that are incompatible with it, but

also claims that are incompatible with it in the context of one's other collateral beliefs—that is, which complete a *set* of claims that are jointly (but perhaps not severally) incompatible with it.

I take it that there is nothing unintelligible about having such practical abilities, fallible and incomplete though they may be, to distinguish claims that are from those that are not contextually incompatible with a given claim. And it is clear that a global updating capacity can be algorithmically elaborated from such abilities to discriminate ranges of counterfactual robustness. But I do not think that this sort of ability is a good candidate for an algorithmic decomposition that is *substantive* in the sense I have given to that term. For I do not see that we can make sense of abilities to discern ranges of counterfactual robustness being exhibited, whether severally or collectively, by *non-discursive* creatures. These are essentially inferential abilities to determine what *would* happen *if*…. They articulate what is already an ability to entertain semantic contents that I will argue in my fifth lecture have a *holistic* dimension in virtue of the fact that sets of them must be taken to be materially incoherent, and to stand in relations of material incompatibility to one another. Being able to entertain semantic contents with that character *is* being sapient. So associating ranges of counterfactual robustness with one's commitments cannot qualify as the endpoint of an algorithmic decomposition of updating abilities that is substantive in the sense required.

It might be objected that this argument depends on endorsing an account of sapience—the avowed target of AI-functionalism—that is entirely too language-centered. Non-linguistic creatures, too, can have beliefs, and do engage in doxastic updating. In response, it must be acknowledged that it is true that the topic I have focused on is language-use, and that my emphasis on deploying *vocabularies* is given aid and comfort in the AI context by the venerable Turing test. But I do not think that the issue of the substantive algorithmic decomposability of the practices-or-abilities that constitute sapience turns on whether we think of the sapience in question as essentially, or only accidentally, linguistic in character. It arises in pretty much the same form no matter where one draws the line. If one defines sapience down (from my point of view) so that orangutans and tigers exhibit it—so that they are

counted as believers, reasoners, and updaters in a full-blooded sense—then the question becomes whether the capacity they are then taken to have to associate ranges of counterfactual robustness with their beliefs is something that can intelligibly be thought to be exhibited by anything that is *not* sapient in the same sense in which they are. If, as I would claim, not, then the argument against the substantive algorithmic decomposability of sapience that proceeds from noting the sensitivity of doxastic updating to the whole context of collateral beliefs goes through for non-linguistic conceptions of sapience as well as for linguistic ones. A problem would arise for the argument only if we could see how to elaborate the ability to associate ranges of counterfactual robustness with the sentences expressing doxastic commitments from the sort of flexible goal-pursuit and obstacle-overcoming abilities exhibited by a skilled and successful non-language-using predator.

What we have to work with in the case of such high-end animals is dispositions: facts about how the animal *would* react *if* things were thus-and-so. Such attitudes *can* underwrite attitudes towards linguistically formulated counterfactuals. They can be what is made explicit by endorsing such conditionals. But I do not think they can underwrite enough of them to fund all the necessary features of the process of doxastic updating of which linguistic creatures as such must be capable. The problem is that the productivity of language[5] guarantees that anything that can talk can form predicates specifying an indefinitely large class of relational properties. As a consequence, any new information about any object carries with it new information of a sort about every other object. For *any* change in *any* property of one changes *some* of the relational properties of *all* the rest. The problem in a nutshell 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, I claim, that *this* ability can be algorithmically decomposed into abilities exhibitable by non-linguistic creatures.

[5] This is a topic I will address in my fifth lecture.

Why not? The logical and computational versions of what the AI community calls the "frame problem" showed that updating requires exercising what turns out to be a crucially important but easily overlooked cognitive skill: the capacity to *ignore* some factors one is capable of attending to. But worrying about the practical engineering problem of how to implement such an ability in finite-state automata revealed a deeper theoretical conceptual problem, which concerns not *how* to ignore some considerations, but *what* to ignore. A simple version of the issue is afforded by the familiar observation that anything is similar to anything else in an infinite number of ways, and also dissimilar to it in an infinite number of ways. For instance, my left little finger and Bach's second Brandenburg concerto are not only different in countless ways, but are similar in that neither is a window-shade nor a prime number, neither existed before 1600 and both can be damaged by careless use of stringed instruments. Dealing with objects as knowers and agents require the ability to privilege some of these respects of similarity and difference—to sort the myriad of such respects into those that are and those that are not relevant to or significant for the inferences, theoretical and practical, to and from the claims about those objects with which one is concerned. In the sort of case I want to focus on, there are lots of complex relational properties that we should usually ignore in our reasoning. Fodor defines a particle as being a 'fridgeon' just in case his fridge is on.[6] So when his fridge turns on, it also turns all the particles in the universe temporarily into fridgeons, and gives every macroscopic physical object the new property of being made of fridgeons. A death in a distant place can give me the new property of having the same eye-color as the oldest living inhabitant of Provo, Utah. Usually I ought to ignore these properties and facts. One of the lessons of the narrower versions of the frame problem is that updating becomes computationally infeasible if I cannot do that, and am accordingly obliged to check every one of my beliefs and the inferences that support them to see whether they are infirmed by those facts—to be sure that my conclusion that the solid floor will bear my weight is

[6] J. A. Fodor, "Modules, Frames, Fridgeons, Sleeping Dogs, and the Music of the Spheres", in Pylyshyn, Z.W. (ed.), *The Robot's Dilemma: The Frame Problem in Artificial Intelligence* [Ablex, Norwood, NJ 1987].

not affected by its suddenly consisting of fridgeons and that my inferential expectation that I will see better if I put on my glasses is still a good one even though my eyes have the new Provo property. For a while there was a small philosophical industry devoted to trying to distinguish what Geach (thinking of McTaggart) called 'Cambridge changes' from real ones.[7] I think we have come to see that this enterprise is a misguided one. For any complex relational property such as being a fridgeon or having old-Provo-colored eyes, we can describe inferential circumstances (however outré) in which the credentials of some significant claim would turn precisely on the presence or absence of that property. What we need to be able to do is not to classify some properties as, in effect, irrelevant *tout court* (irrelevant to what?), but for each inference to distinguish the considerations that are irrelevant to its goodness, which should accordingly be ignored. This ability is necessary to deal with what Fodor calls epistemological 'isotropy': the fact that any belief is potentially evidentially relevant to any other, given a suitable context of collateral beliefs.[8]

I am claiming that:

- 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[9], and for many complex relational properties, only those with access to the combinatorial

[7] Peter Geach, *Logic Matters* [Berkeley, CA; University of California Press, 1980] p. 321.
[8] J. A. Fodor, J.A., *The Mind Doesn't Work That Way* [MIT Press, Cambridge, MA 2000], Chapter 2.
[9] A PP-necessity claim. Compare Sellars's argument about 'looks' talk, which turns on the claim that one cannot withhold an endorsement one is not capable of giving.

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.

This last claim is a somewhat delicate one. I am *not* using as a premise the claim that we cannot make sense of the possibility of substantively algorithmically decomposing the capacity to be aware of a full range of complex relational properties, by deploying a suitable vocabulary. That is part of the conclusion I am arguing for. I *am* claiming, first, that the ability to ignore the vast majority of complex relational properties that are irrelevant to a given inference in the sense that they fall within its range of counterfactual robustness cannot be taken as *primitive* with respect to a *substantive* algorithmic practical decomposition of discursive practices-or-abilities, and, second, that we have no idea at all how even primitive non-discursive abilities that *could* be substantively algorithmically elaborated into the capacity to form the complex predicates in question could be further elaborated so as to permit the sorting of them into those that do and those that do not belong in the range of counterfactual robustness of a particular inference.

Because inferences link different concepts, taking any inferential connections to be semantically essential to the concepts involved results in some kind of semantic holism.[10] Still, some aspects of inferential roles are fully compositional. In particular, *substitution* inferences involving singular terms and predicates are. These are inferences commitment to which can be made explicit by endorsing identity claims and quantified conditionals, such as "Robert Hooke is (was, =) the inventor of the universal joint" and "Any person who sings, breathes." But what is at issue here is the capacity to assess the goodness of *non-monotonic* material inferences, by practically associating with each a range of counterfactual robustness: responsively distinguishing the further collateral premises that would not infirm a given good material inference. This function is not computable by substitution inferences, or even from the incompatibilities of sets of non-logical sentences. And that lack cannot be repaired by introducing a non-monotonic *logic*, because what is at issue is the propriety of non-monotonic *material* inferences.[11]

The best arguments I can come up with against the algorithmic practical elaboration version of AI-functionalism are arguments against the substantive algorithmic decomposability of the practical capacity to associate a range of counterfactual robustness with material inferences involving complex relational properties. The issue is an empirical one, and these arguments are accordingly probative rather than dispositive. They offer reasons supporting pessimism about the algorithmic practical elaboration AI thesis, but not reasons obliging us to offer a negative verdict.

Section 5: Practical Elaboration by Training

One might reasonably wonder whether, if the sort of argument I have sketched against the substantive algorithmic decomposability of autonomous discursive practices were successful, it would not prove too much. Nonlinguistic creatures do, after all, acquire the ability to engage in discursive practices. They do cross the boundary I have been worrying about, and begin deploying vocabularies. This is true both of human infants and, at some point in the past, of our hominid ancestors. The ontogenetic and phylogenetic acquisition of discursive capacities did not

[10] This is a claim about a kind of semantic holism that goes beyond that to be discussed in the fifth lecture.
[11] This point is discussed in more detail in the next two lectures.

and does not happen by magic. (And at least in this context, I do not think the image of "the light dawning slowly over the whole" is of much help.) If discursive practices-or-abilities really are not substantively algorithmically decomposable without remainder into non-discursive ones, how *are* we to understand the development of discursive out of non-discursive practices?

I think the answer is that apart from algorithmic elaboration there is another, more basic sort of PP-sufficiency relation—another way in which one set of practices-or-abilities can practically suffice for the acquisition of another. Sometimes those who can engage in one set of practices can learn or be trained to engage in another—not because the target practices can be algorithmically elaborated from the original ones, or from some further set into which they can be decomposed, but just because as a matter of contingent empirical fact concerning creatures of that particular kind, anyone who has the one set of capacities can be brought to have the other as well. So it might be that those who can draw realistic portraits of horses can be brought also to draw realistic portraits of humans, forge signatures, fold origami gracefully, and arrange flowers. If so, no doubt our account of why these other abilities were especially accessible to those who possess the original one would invoke something like "eye-hand co-ordination" or "fine-muscle control". But that is not at all to say that there must be some set of specifiable basic abilities out of which, say, the capacity to draw a good likeness of a friend could be algorithmically elaborated. That capacity might admit of no algorithmic decomposition. Certainly the fact that people who can do some other sorts of things can learn or be taught also to do this does not entail or require that there be such a decomposition.

For reasons that will become clear only a little further along, I will be treating *learning* as a special case of *training*—corresponding to the somewhat degenerate case of a trainer who is, as it were, blind or stupid—rather

than, as is usual, the other way around. And I will discuss training in the context of a somewhat regimented distinction between *teaching* and *training*. Both are ways of eliciting new abilities. But on this usage, teaching someone to do something does that by *telling* the student what to do and how to do it, while training someone to do something does that by *showing* the trainee what to do and how to do it, guiding him through a course of practice exercises whose results are critically assessed. In this sense, adult beavers may train their pups to build dams, but cannot teach them to do it, and no amount of teaching can bring someone to be able to ride a bicycle or drive a car. Long division can indeed be taught to those who already know how to multiply and subtract, but one can only be trained to remember the names and order of the letters of the English alphabet. That a certain amount of rote repetition suffices to train five-year-olds to have that latter ability is a more-or-less brute fact about us. And if one could not be trained to do things like that, and to recognize colors and shapes and body parts and so on, one could not be taught to do anything. The abilities that are primitive with respect to an algorithmic decomposition of some complex ability are themselves typically the result of *developing* other abilities by training those who exhibit them.

When as a matter of 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, I will say that the second can be "practically elaborated by training" from the first. Like algorithmic elaboration, practical elaboration by training is a kind of PP-sufficiency relation. The hallmark of the difference between them is that we can say exactly and in advance what the practices that *implement* PP-sufficiency by algorithmic elaboration are—what *else* besides the primitive abilities one must be able to do in order to elaborate them algorithmically into the target ability. These elaborative abilities are things like response substitution and state formation, and in general the abilities that suffice to execute a conditional branched-schedule algorithm. These algorithmic elaborative abilities are all that is needed to turn the capacity to multiply and subtract into the capacity to do decimal division. By contrast, in the case of practical elaboration by training, we have no idea how to specify in advance the abilities that implement the sufficiency of rote repetition for

memorizing the alphabet, or practice in catching a ball or drawing a face. And where we *can* say something about the abilities that implement PP-sufficiency relations of the practical-elaboration-by-training sort, we find that they both vary wildly from case to case, and depend heavily on parochial biological, psychological, sociological, and historical contingencies. Finally, where the question of whether one set of well-defined practices-or-abilities can be elaborated algorithmically into another is one that can in principle be settled *a priori*, from one's armchair, the question of whether it is practically PP-sufficient for some particular creature or kind of creature, in a particular context, by some training regimen is one that can only be settled empirically.

I think an appreciation of the centrality of this sort of PP-sufficiency relation—which obtains when as a matter of fact creatures of a certain sort who can engage in a practice (exhibit an ability) can be brought or can learn to engage in (or exhibit) another—is one of the master ideas animating the thought of the later Wittgenstein. Again and again he emphasizes the extent to which our discursive practices are made possible by the fact that as a matter of contingent fact, those who have one set of abilities or can engage in one set of practices can be brought by training to exhibit or engage in another. We can be trained to count, associate sounds with written shapes, and respond to sign-posts, and to exercise those ability in new cases by "going on in the same way" as others who share our training (and wiring) would. Wittgenstein is of course concerned to show us to what extent and in how many ways our discursive practices-or-abilities depend on things that we could not be *taught* to do (by being told) if we could not be *trained* to do them (by being shown). But I think he also sees practical elaboration by training as the principal motor of our discursive practices-or-abilities, as what gives them their theoretically

motley but practically tractable shapes. As I read him, Wittgenstein thinks that *the* most fundamental discursive phenomenon is this way in which the abilities required to deploy one vocabulary can be practically *extended*, elaborated, or developed so as to constitute the ability to deploy some further vocabulary. We may think in this connection of the examples I mentioned in my first lecture, of the sort of thought-experiments he invites us to conduct concerning this sort of process of *pragmatic projection* of one practice into another: the fact that people who could already use proper names for people could catch on to the practice of using them also for rivers, and that people who could already talk about having gold in their teeth could catch on to talking about having pains in their teeth. The way in which prior abilities are recruited by training in the service of developing new ones is in general unsystematic, not codifiable in rules or algorithms, and not predictable or explicable from first principles. Wittgenstein sees this sort of non-algorithmic practical elaboration as ubiquitous and pervasive. Further, much of the discursive training is not intentional and goal-directed at an already extant and characterizable practice-or-ability, as elementary education is. Rather, it is the spontaneous, contingent, unforeseeable result of deploying familiar vocabulary in novel circumstances. It results in a permanent process of practical discursive mutation that is on the one hand mediated by the productivity of language, and on the other limits its diachronic systematicity.[12]

So the answer to the question with which I began this section is that 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

[12] Mark Wilson's deep and important book *Wandering Significance: An Essay On Conceptual Behaviour* [Oxford University Press, 2006] offers a subtle and compelling account of just how various and uncodifiable the practical capacities that underlie our conceptual classifications are, and how the strategies implicitly employed to cobble together a unified conceptual façade out of the diversity that results from extending the application of our terms to unforeseen realms resist systematization of the sort that seems to be demanded by standard semantic theories.

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. 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 this sort of elaboration 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.

I said at the outset of my story that one of the aims of the sort of analytical pragmatism for which I am seeking to sketch a theoretical basis is to show how Wittgenstein's pragmatist insights need not be taken to underwrite a theoretical quietism antithetical to the project of traditional philosophical analysis, but how those insights can instead be taken on board and pressed into the service of a further pragmatic development and elaboration of that project. Acknowledging the pervasiveness and centrality of non-algorithmic practical elaboration by training need not be the death of theoretical analysis of discursive practice and its relation to the semantic contents expressed by the vocabularies deployed in that practice. 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. I would like to close this lecture by outlining one analytic issue that I think is raised directly by the consideration of what I will call *pedagogic* practical elaboration and decomposition.

I have pointed out that one set of abilities can be elaborated into another by a process of *training*, rather than algorithmically, and that the practices-or-abilities that implement algorithmic elaboration are neither necessary nor sufficient for this sort of practical elaboration. Besides this negative characterization, what can we say positively about what training is? Most generally, I think of training as a course of *experience*, in Hegel's and Dewey's sense (Erfahrung rather than Erlebnis) of a feedback loop of perception, responsive performance, and perception of the results of the performance. When we think about the practices-or-abilities that *implement* elaboration-by-training, we can think about them both on the side of the trainer and of the trainee (though both learning—training without a trainer—and self-training, which is not the same thing, are also important species). A course of training implements a *pedagogic* elaboration of one set of abilities into another. We can think of it very abstractly as having as its basic unit a stimulus (perhaps provided by the trainer), a response on the part of the trainee, a response by the trainer to that response, and a response to that response by the trainee that alters his dispositions to respond to future stimuli. A constellation of such units constitutes a *course* of *training*. The sequence of stimuli and responses-to-responses may be fixed and rigid—corresponding in this domain to a straight-schedule algorithm. Or it may be flexible, in that not only the trainer's responses to the trainee's responses vary with the trainee's responses, but also the sequence of presented stimuli. Such a course of training corresponds to a conditional branched-schedule algorithm. But training need not be codifiable as a training *regimen* in either way—as for instance most "on the job" training is not, and much learning in childhood that takes place outside of schools. Such training exploits contingent, adventitious stimuli, in a haphazard sequence. At least one important species of such training can be modeled by connectionist networks, which develop one set of primitive abilities into a set of target abilities by training. The training may be guided by an explicit representation, on the part of the trainer, of the abilities aimed at. Or it may be something much more implicit, where the trainer can distinguish between (respond differentially to) correct and incorrect, more and less skilled performances, without being able to say what the difference consists in. And the course of training may be more or less explicit in the mind of the trainer. At the

limit, the algorithm corresponding to the flexible course of training may be formulable by the trainer. Even where the training regimen can be made explicit by the trainer, the result need not be teaching rather than training. For as I am using the terms, teaching is training by *describing* what it is one is trying to do and how to go about doing it. It is a way of getting one to *do* something by *saying* what one is to do. Where training in general is a matter of implementing PP-sufficiency, teaching does so by a VP-sufficient characterization or specification of what one must do in order to turn an initial set of practices-or-abilities into the targeted ones.

Section 6: Algorithmic Pedagogical Decomposition and Pedagogical Politics

I suggested that what in a course of training is most analogous to algorithmic elaboration of abilities is *pedagogic* elaboration in the form of a training regimen. The differences must be kept firmly in mind. What we may call an *executive* algorithmic elaboration is *logically* sufficient to yield the target ability. If one multiplies and subtracts, performing each operation on the results of the other's calculations according to the right algorithm, the result is guaranteed to be a decimal representation of the quotient of the two numbers one operated on. By contrast, that a certain kind of rote repetition exercise suffices for most children to learn the alphabet or the single-digit multiplication table is a matter of *psychology*—or, to use a fine, old-fashioned phrase, of the *natural history* of humans—not of logic. Even where the target ability is algorithmically decomposable into available primitive abilities—as for instance subtraction is to counting—it not only does not follow that one can train someone by *explaining* the algorithm (that is, by *telling* about it, *teaching* it), but thinking about the algorithm may be of very little use to the trainer in solving the *pedagogical* problem of how as a matter of fact it is possible to develop the primitive abilities into the target ability.

In rare but important cases in early education, we have *completely solved* the problem of how to pedagogically elaborate one set of abilities into another. What it means 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 TOTE cycle of responses to their responses *in fact* (though without the guarantee of any principle) get them to catch on to the target ability. Training pupils who can already *count* to be able to *add* is essentially a solved pedagogical problem in this sense. That is, starting with pupils of widely varying abilities and prior experiences, who share only the prior ability to count, there is a flowchart of differentially elicited instructions, tests, and exercises that will lead all of them to the target skill of being able correctly to add pairs of arbitrary multidigit numbers. A common initial lesson or exercise is followed by a diagnostic test. The results of that test then determine, for each pupil, which of an array of possible second lessons or exercises is appropriate, followed by further tests whose results are interpreted as differentially calling for different exercises, and so on. This flowchart determines a Test-Operate-Test-Exit cycle of training that incorporates a *pedagogic* (as opposed to an executive) *algorithm*. In some cases, one can think of the intermediate tests as demonstrating the presence or absence of characterizable sub-skills, each of which is necessary and all of which are sufficient for the ultimate target skill, and which stand in various relations of presupposition to one another. But it need not be possible to isolate such a set of sub-skills, and it is important to realize that where it is, they are not in general *algorithmically* sufficient for addition in the executive sense, and cannot be derived by any amount of thinking about *addition* by itself. For the practical PP-sufficiencies and necessities exploited by a pedagogical algorithm depend on and exploit contingent empirical *psychological* (and sociological) particularities of the subjects and settings.

I am told by those who know about these things that teaching *multiplication* to pupils who can add and subtract is also in this sense a completely solved pedagogical problem, but that in spite of massive investigative efforts to date, *subtraction* remains an essentially unsolved pedagogical problem, and *division*, in the form of mastery of fractions, a tantalizing, so far

intractable pedagogical mystery.In the absence of a practical pedagogical algorithm, those charged with eliciting and developing such skills must fall back on rougher heuristics and the sort of practical know-how gleaned from many years of trial-and-error training of a wide variety of candidates. Many of the mentors *learn* how to do it pretty well, and we may even be able to *train* them pretty well to be able to do it, but in the absence of a complete pedagogical solution to the practical training problem, we cannot *teach* them how to do this sort of training. For only a practical algorithmic decomposition of the *training* practices yields a vocabulary that is VP-sufficient to specify those PP-sufficient training practices.

Incompletely solved pedagogical problems—not just in specialized cases in elementary education, but at all levels and across the board—raise a broad issue of institutional *politics*, which seems to me to penetrate deeply into the society as a whole. The pedagogical algorithm that constitutes a complete solution to a training problem empirically guarantees that for every candidate who exhibits the basic input skills, there is a training regimen that implements the practical PP-sufficiency of those skills for the target skills: a course of experience and practice that will elicit those skills from or confer them upon that trainee. Augustine marveled at the rumored ability of a monk to gather the sense of a text without pronouncing aloud the words on the page and then listening to them (and rode three days on a mule to test it), and one of Samuel Pepys' distinctive qualifications for his positions as Secretary of the Admiralty was his mastery of the arithmetic required for double-entry bookkeeping. Today we take it for granted that we can train almost everyone to read silently and to add up long columns of figures. But for abilities for which the pedagogic problem has not been completely solved, where we do not yet have an algorithmic decomposition of the practical training process, candidates who exhibit all the relevant primitive abilities are *de facto* sorted by the training regimens we do have not only by the number of iterations of the Test-Operate-Test loop it takes for them to acquire the target ability, but by whether they can be brought to that point at all.

Matter-of-factual PP-necessity relations among practices-or-abilities require that the outputs of some training regimens serve as the inputs to others—that some of the abilities treated as primitive (as practically *a priori*) by the one are achievable only as the target abilities (the practical *a posteriori*) of others. It follows that the effects of failing to acquire one ability—falling into the missing, incompletely mapped portion of an ideally complete pedagogical solution of which some actual training regimen is a mere fragment—will be strongly cumulative within a sequence of courses of learning-and-training.

The broadly political issue I want to point to concerns how, in the context of these very general considerations, we should think about one element of just treatment of individuals by institutions, whether they be schools, corporations, governments, or the society as a whole. We might, as a demand of justice or simply a counsel of social engineering, want some rewards to be proportioned to productive achievements according to some definition of the latter. Among the crucial necessary conditions of any such achievement is the possession of certain skills or abilities—along with such other conditions as the opportunity to exercise those skills or abilities, the resources required to do so, the effort one expends in doing so, and perhaps also a residue of good fortune concerning the actual consequences of one's efforts. Each of these kinds of condition of achievement raises potential issues for assessment of the justice (and, for that matter, of the engineering efficacy) of the reward scheme: equality of opportunity and of access to resources, the relative weights to be accorded to hard work that unforeseeably turns out not to yield the desired result as opposed to good outcomes accidentally achieved, and so on. But the one on which I want to focus concerns the possession of the skills or abilities whose exercise is conditioned or qualified in all those other ways. It seems that 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.

One attitude is that 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. What are being assessed here are the practical-elaborative abilities that *implement* the PP-sufficiency of some set of primitive abilities for the target ability, in the context of the course of experience yielded by a training regimen. 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.

The view that forms the opposite pole of the dimension I am pointing to focuses on the relativity of the hierarchical sorting of candidates into more or less trainable 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—just being more trainable, 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 a *practical* commitment to find and implement those solutions. It is an extreme, indeed utopian, pedagogical egalitarianism.

Taken to the limit, the pedagogic egalitarian view may seem to rest on a literally unbelievable premise: that whatever *some* human can (learn or be trained to) do, *any* human can (learn or be trained to) do. And the evaluative component implicit in the cognitive commitment as I stated may seem no more plausible: that there is something wrong with rewarding any ability of which that claim is *not* true. A more defensible version of pedagogic egalitarianism results if the latter commitment is softened so as to claim merely that special arguments must be given in each case for the valorization of differences in ability for which we have found no complete pedagogic solution. The first element of the cognitive component can correspondingly be interpreted in terms of the practical commitment so as to claim that we have no right to assume, for any given skill or ability for which we have as yet no complete pedagogic solution, that that is because there *is* in principle no such solution.

The empirical-hierarchical attitude is *conservative* in treating extant institutionalized training regimens as given and fixed, and the utopian pedagogical egalitarian attitude is *progressive* in its commitment to transform them. As political attitudes, they articulate one

understanding of the nature/nurture aspect of traditional right/left alignments. Between them lies a whole array of more nuanced principles for assigning reciprocal, co-ordinate responsibility to training or trainers, on the one hand, and trainees on the other. We need not simply choose between the strategies of holding actual training regimens fixed and hierarchically sorting humans with respect to them, on the one hand, and holding the actual practical-elaborative abilities of humans fixed and sorting training regimens with respect to them, on the other.

But my purpose in gesturing at this issue of pedagogical politics here has not been to recommend one or another way of approaching it. Assessing the plausibility of the broadened, practical version of the thesis of artificial intelligence led to the notion of practical PP-sufficiency by training. My aim in this final section has been to lay alongside the postulate of universal practical *executive* algorithmic decomposability of discursive abilities, characteristic of AI, the postulate of universal practical *pedagogic* algorithmic decomposability of discursive abilities characteristic of utopian pedagogic egalitarianism, and to point to an issue of considerable philosophical, cultural, and political significance that it raises. As a result, the argument of the lecture as a whole has described a narrative arc taking us from Turing, through Wittgenstein, to Dewey.

My last three lectures will address modal vocabulary, normative vocabulary, and the pragmatically mediated semantic relations they stand in to ordinary objective, empirical, and naturalistic vocabularies, and to each other. I will argue that both the deontic vocabulary of conceptual norms and the alethic vocabulary of laws and possibilities can be elaborated from and are explicative of features necessarily exhibited by any autonomous discursive practice.

Thinking about the pragmatically mediated semantic relations they stand in to each other turns out to provide a new way of understanding the subjective and objective poles of the intentional nexus of knowers-and-agents with their world. Along the way, I will show how normative vocabulary can serve as an expressively bootstrapping pragmatic metavocabulary for modal vocabulary, and, in the fifth lecture, how that fact makes possible a new sort of formal semantics for logical and modal vocabulary, as well as for ordinary empirical descriptive vocabulary.

END

[(4-10 a: 8499 words in large type (=28.33 pages)]