Chapter 1
Varieties of Nativism

Chomsky's Nativism

The present phase of nativistic theorizing about the cognitive mind began with two suggestions of Noam Chomsky's: that there are substantive, universal constraints on the kinds of grammars that natural languages can have; and that these constraints express correspondingly substantive and universal properties of human psychology (determined, presumably, by the characteristic genetic endowment of our species). In effect, Chomsky predicted the convergence of two lines of research:

- On the one hand, empirical investigation of the range of grammatical structures that human languages exhibit would estimate the limits within which it is possible for them to vary. One then subtracts the ways that human languages can differ from the ways in which it is conceivable that languages *could* differ. The remainder after the subtraction is the set of linguistic universals that implicitly define "possible human language."[1]
- On the other hand, empirical investigations of the conditions under which children learn to talk would estimate the information their linguistic environments provide, hence how much poverty of the stimulus the language learning process tolerates. One then subtracts the information that is in the environment from the information that is required for the child to achieve linguistic mastery. The remainder after the subtraction is what the child's innate knowledge contributes to the language acquisition process.

If everything goes well, it should turn out that what the child innately knows will be the same universal principles that constrain

the humanly possible languages. Such a convergence would explain, in one stroke, both why human languages don't differ *arbitrarily* and also why (pace occasional sentimental claims on behalf of dolphins and chimpanzees) only human beings seem to be any good at learning them.

In principle, the research strategy that Chomsky proposed seems perfectly straightforward to execute. One need only determine the empirical values of the relevant parameters, perform the indicated subtractions, and then compare the remainders. So why, you might wonder, didn't somebody just get a grant and do it? In practice that turned out not to be easy. For one thing, it's not easy for cognitive scientists to get grants if they are working on questions of any theoretical interest. (To ensure this is a main function of the institution of peer review.) And, for another thing, even rational people can disagree about how much, and in what ways, languages actually differ; and about whether the residual similarities might after all be "explained away" without resort to nativistic postulations (perhaps by appealing to historical or environmental factors, or to the functional properties that any language would need to have if it is to be expressive and efficient). Likewise, it is no small matter to figure out what information the child's linguistic environment makes available to the acquisition process; or how much of what it makes available the child actually exploits; or how much of what the child actually exploits he could have done without, consonant with achieving normal fluency by the normal means. One can't, of course, perform Kasper Hauser experiments on the offspring of one's conspecifics.

So the argument that Chomsky started all those years ago continues unabated. I assume its general outlines are familiar, and I won't rehearse them further here. What's most striking for our purposes is a point about his view that Chomsky has himself often emphasized: Insofar as it concerns the relation between human language and human nature, his position is continuous with—indeed, practically indistinguishable from—one that philosophical rationalists have defended for centuries. Except for the characteristically modern identification of "human nature"

with "what the human genotype specifies," Chomsky's ideas about innateness would have been intelligible to Plato; and they would have been intelligible in much the terms of the present debate.

This is because Chomsky's nativism is primarily a thesis about knowledge and belief; it aligns problems in the theory of language with those in the theory of knowledge. Indeed, as often as not, the vocabulary in which Chomsky frames linguistic issues is explicitly epistemological. Thus, the grammar of a language specifies what its speaker/hearers have to *know* qua speakers and hearers; and the goal of the child's language acquisition process is to construct a *theory* of the language that correctly expresses this grammatical knowledge. Likewise, the central problem of language acquisition arises from the poverty of the "primary linguistic *data*" from which the child effects this construction; and the proposed solution of the problem is that much of the knowledge that linguistic competence depends on is available to the child *a priori* (i.e., prior to learning). Everything I've put in italics belongs to the epistemologist's vocabulary; it is, to repeat, primarily epistemological nativism that Chomsky shares with the rationalists. When Plato asks what the slave boy knows about geometry, and where on earth he could have learned it, it really is much the same question that Chomsky asks about what speaker/hearers know about their language and where on earth they could have learned that. There is, I think, no equivocation on the key terms.[2]

By contrast, New Synthesis psychological theories of the kind that Pinker and Plotkin espouse are typically about not *epistemic states* but *cognitive processes;* for example, the mental processes involved in thinking, learning, and perceiving. The key idea of New Synthesis psychology is that cognitive processes are *computational;* and the notion of computation thus appealed to borrows heavily from the foundational work of Alan Turing. A computation, according to this understanding, is a formal operation on syntactically structured representations. Accordingly, a mental process, qua computation, is a formal operation on syntactically structured mental representations. We'll return to this idea quite

soon and at length. Suffice it, for the moment, that whereas Chomsky's rationalism consists primarily in nativism about the knowledge that cognitive capacities manifest, New Synthesis rationalism consists primarily in nativism about the computational mechanisms that exploit such knowledge for the purposes of cognition. To put it in a nutshell: *What's new about the New Synthesis is mostly the consequence of conjoining a rationalist epistemology with a syntactic notion of mental computation.*

The attempt to ground psychology in the idea that mental processes are computations is a main topic of the discussion to follow. I'm mostly interested in telling you what I think is right about this idea and what I think isn't. But first I have to tell you how it's supposed to work. This will take some fairly extended exegesis. Please do bear with me. Unlike epistemic nativism, computational nativism really is a new kind of rationalist theory; whereas Plato would have understood Chomsky well enough, I doubt that he would have understood Turing at all.

The New Synthesis

1. Computation
It's a remarkable fact that you can tell, just by looking at it, that any (declarative) sentence of the syntactic form *P* and *Q* ("John swims and Mary drinks," for example) is true if and only if *P* and *Q* are themselves both true; that is, that sentences of the form *P and Q* entail, and are entailed by, the corresponding sentences *P*, *Q*. To say that "you can tell this just by looking" is to claim that you don't have to know anything about what either *P* or *Q means* to see that these entailment relations hold, and that you also don't have to know anything about the nonlinguistic world.[3] This really is remarkable since, after all, it's what they mean, together with the facts about the nonlinguistic world, that decide whether *P* or *Q are* true.

This line of thought is often summarized by saying that some inferences are **"**formally valid," which is in turn to say that they hold just in virtue of the "syntax" of the sentences that enter into

them.[4] It was Turing's great discovery that machines can be designed to evaluate any inference that is formally valid in that sense. That's because, although machines are awful at figuring out what things mean and aren't much better at figuring out what's going on in the world, you can build them so that they are quite good at detecting and responding to syntactic properties and relations. That, in turn, is because the syntax of a sentence reduces to the identity and arrangement of its elementary parts, and, at least in the artificial languages that machines compute in, these elementary parts and arrangements can be exhaustively itemized, and the machine specifically designed to detect them.

So: Turing showed us how to make a computing machine that will recognize any argument that is valid in virtue of its syntax; and the basic thesis of the new psychological synthesis is that cognitive mental processes are (perhaps exhaustively) constituted by the kinds of operations that such machines perform.

Notice, in particular, that the reliance on syntax is essential; it's *only* if the sufficient conditions for an inference to be truth preserving are syntactic that Turing guarantees that a machine is able to recognize its validity. So if, like New Synthesis theorists, you propose to co-opt Turing's account of the nature of computation for use in a cognitive psychology of thought, you will have to assume that *thoughts themselves have syntactic structure.* What's on offer at the price of this assumption is the prospect of a theory that explains how, in a variety of kinds of cases, mental processes can lead, reliably, from one true thought to another. That sounds to me like a bargain.[5]

Right; so much, for now, for Turing's account of computation. What has all this got to do with the rationalist tradition in psychology?

The New Synthesis Continued

2. Rationalist psychology

Rationalists are nativists practically by definition; by contrast, the rationalist consensus about the nature of mental processes is

less than transparent to first impressions. Still, I think there is such a consensus, epitomized perhaps by Kant; and that it has its roots in Aristotle and reaches us via such of the Scholastics as William of Occam. If this were a work of scholarship, and if I were a scholar, I'd try to make some sort of case for these historical claims; but it's not, and I'm not, so I won't. Suffice it to make explicit what I take the main idea of rationalist psychology to be, and how I suppose that it connects with the Turing-style account of computation sketched above.

The main idea of rationalist psychology is that beliefs, desires, thoughts, and the like have logical forms, and that their logical forms are among the determinants of the roles they play in mental processes. For example, *John swims and Mary drinks* is a *conjunctive* belief, and that is why having it can lead one to infer that John swims; *there aren't any unicorns* is a *negative existential* belief, and that is why having it can lead one to the infer that Alfred is not a unicorn. And so forth. Accordingly, I will use the term "rationalist psychology" for any theory according to which (at least some) mental states have logical form, and the causal role of a mental state depends (at least inter alia) on what logical form it has.[6]

What follows is a number of exegetical comments on the general character of rationalist psychologies so construed, and on why they accommodate themselves naturally to the thesis that mental processes are computations. We'll see that what connects the two is primarily the idea that the logical form of a thought might be reconstructed by the syntax of a mental representation that expresses it.

Comments (in no particular order):

> • Beliefs, desires, thoughts, and the like[7] (from here on, I'll call them all "propositional attitudes") have their logical forms *intrinsically.* Which is to say not only that if *x* and *y* are propositional attitudes of different logical forms they are ipso facto different mental particulars, but also that they are ipso facto mental particulars of different types. Sam's

belief that $P \lor Q$, for example, is ipso facto of a different type than his belief that $\sim(\sim P \& \sim Q)$, even though the two are, of course, logically equivalent.

• Propositional attitudes with different *contents* may have the same logical form. The belief that there isn't any Santa Clause has the same logical form as the belief that there aren't any unicorns even though they are, of course, different beliefs.

• Assume, for simplicity of exposition, that the paradigmatic propositional attitude is a belief that a certain individual has a certain property, for example, *that John is bald.* Such a belief has the logical form *Fa*, where "*F*" expresses the property that the individual is believed to have (e.g., being bald) and "*a*" specifies the individual that is believed to have that property (e.g., John). A belief of the form *Fa* is true if and only if the individual in question actually does have the property in question.

• As in the preceding example, so too in the general case: Propositional attitudes are complex objects; propositional attitudes have parts. In what follows, I'll often refer to the parts of a propositional attitude as its "constituents." The constituents of the belief that John is bald include: the part that expresses the property of being bald and the part that specifies John. In the psychologist's usage, the constituents of propositional attitudes are often called "concepts."[8]

• The logical form of a propositional attitude is not (repeat: is *not)* reducible to the causal relations among its constituents (which is not to deny that it may be reducible to some causal relations or other). *This is a fundamental difference between rationalist and empiricist psychologies:* whereas, according to the latter, the structure of a thought is fully determined by specifying the pattern of associations among its constituents, according to the former, it is an independent parameter.[9] It is basically because rationalists distinguish between the structure of a thought and what is sometimes called its degree of "associative integration" that they can explain how it is possible to come to believe

the very same thing that one used to doubt or deny (or vice versa.)

I want to be as clear as I can about this, since I take it to be what primarily distinguishes computational psychology from the (connectionistic) associationism that is the main current alternative. Suppose I only *sort* of think that John is bald, whereas you are utterly certain that he is. Suppose, moreover, that it really matters to you whether John is bald, whereas I don't actually much care. In that case, your thinking *John* might cause you to think *bald* (or *he's bald)* with absolutely mechanical regularity, whereas my thinking *John* might cause me to think *bald* at most only now and then, or even not at all. Still, according to the present view, your thought that John is bald is a propositional attitude of exactly the same type as mine, and so a fortiori, they have the same logical form. So, to repeat, its logical form and the causal relations that may hold among its constituents are independent parameters of a propositional attitude according to rationalist psychologies.[10]

• Suppose it's right that mental states can have logical forms to which mental processes are sensitive. The question remains *how* logical forms could determine causal powers. I'm not enough of a historian to know whether the tradition of philosophical rationalism had a consensus view on this question. But it wouldn't surprise me much to hear that it didn't, since rationalists have generally been wary of thinking of mental processes as *causal* at all.[11] It was sufficient to their purposes simply to insist, as I have also done, that the logical form of a thought isn't constituted by the causal relations among its constituents; a fortiori, it isn't constituted by the *associative* relations among its constituents.

But, of course, cognitive scientists generally *do* want to think of mental processes as causal. So if they wish to co-opt the rationalist idea that thoughts have their role in mental processes in virtue of, inter alia, their logical forms, they have to have a view about how logical form could determine causal powers. Just *saying* it does isn't good enough;

you need a mechanism. Conjoining Turing's kind of RTM to a rationalist psychology is what's supposed to provide it: For each propositional attitude that has a causal role in a mental life, there's a corresponding mental representation. Mental representations are concrete particulars, and so are allowed to cause things to happen. Also, mental representations have syntactic structures, to which mental processes are sensitive qua computations. *And the logical form of a propositional attitude supervenes on the syntax of the mental representation that corresponds to it.*[12] That is, disjunctive propositional attitudes (i.e., attitudes whose *logical* form is disjunctive) correspond to disjunctive mental representations (i.e., to mental representations whose syntactic form is disjunctive); conjunctive propositional attitudes correspond to mental representations whose *syntactic* form is conjunctive; existentially quantified propositional attitudes correspond to mental representations whose syntax is existentially quantified . . . and so on for every case in which the logical form of an attitude is invoked to explain its role in mental life.[13]

Perhaps now it starts to be clear why the notion of computation plays such a central role in how rationalist cognitive scientists think about the mind these days. A psychology (rationalist, empiricist, or whatever) needs to do more than just enunciate the laws it claims that mental processes obey. It also needs to explain *what kind of thing a mind could be* such that those laws are true of it; which is once again to say that it needs to specify a mechanism. Empiricists hold, more or less explicitly, that typical psychological laws are generalizations that specify how causal relations among mental states alter as a function of a creature's experience. Associationism provided empiricists with an explanation of why such generalizations hold, namely, that they are all special cases of the associative laws, which are themselves presumed to be innate.[14] By contrast, a rationalist psychology says that typical laws about the mind specify ways in which the logical form of a

mental state determines its role in mental processes. So a rationalist is in need of a theory about how a mental process *could* be sensitive to the logical form of mental states. This theory can't, of course, be associationistic, since associative relations among mental states are supposed to hold *not* in virtue of logical form, but rather in virtue of statistical facts about (e.g.) how often they have occurred together, or how often their occurring together has lead to reinforcement, etc. Turing's notion of computation provides exactly what a rationalist cognitive scientist needs to fill this gap: It does for rationalists what the laws of association would have done for empiricists if only associationism had been true.

• Finally, it's prima facie plausible that computations in Turing's sense should somehow be what implement rationalist psychological theories. For, just as being truth preserving is the characteristic virtue of *computations* as Turing understands them, so too it is the characteristic virtue of *mental processes* as rationalists understand them. One true thought tends to lead to another in the course of cognition, and it is among the great mysteries about the mind how this could be so. Maybe this mystery can be explained on the assumption that typical inferences, insofar as they are valid in virtue of the logical structure of the thoughts involved, are implemented by computations that are driven by the syntactic structure of the corresponding mental representations.[15]

Hence a provisional merger between rationalist psychology and Turing's account of computation, of which the following are the main principles:

The Computational Theory of Mind (= a rationalist psychology implemented by syntactic processes)
i. Thoughts have their causal roles in virtue of, inter alia, their logical form.
ii. The logical form of a thought supervenes on the syntactic form of the corresponding mental representation.

iii. Mental processes (including, paradigmatically, think-ing) are computations, that is, they are operations defined on the syntax of mental representations, and they are reli-ably truth preserving in indefinitely many cases.

The prima facie virtues of effecting this merger is that it (maybe) allows us to solve the two central problems of rationalist psy-chology mentioned above: "What determines the logical form of a thought?" and "How does the logical form of a thought deter-mine its causal powers?" Answer: The logical form of a thought supervenes on the syntax of the corresponding mental represen-tation,[16] and the logical form of a thought determines its causal powers because the syntax of a mental representation determines its computational role, as per the operations of Turing machines. So we can now (maybe) explain how thinking could be both rational and mechanical. Thinking can be rational because syn-tactically specified operations can be truth preserving insofar as they reconstruct relations of logical form; thinking can be mechanical because Turing machines are machines.[17]

However things eventually work out for computational nativism in cognitive science, this really is a lovely idea and we should pause a moment to admire it. Rationality is a normative property; that is, it's one that a mental process *ought* to have. This is the first time that there has ever been a remotely plausible mechanical theory of the causal powers of a normative property. The first time ever.

We now have about half of the New Synthesis in place: The cognitive mind contains whatever innate content "poverty of the stimulus" arguments require it to contain, together with an innate Turing architecture of syntactically structured mental rep-resentations and syntactically driven computational operations defined on these representations. The New Synthesis thus shares with traditional rationalism its emphasis on innate content; but it has added Turing's idea that mental architecture is computa-tional in the proprietary syntactic sense. To round off this expo-sition of computational nativism, we need to explain why New Synthesis psychologists are so often proponents of the thesis that

cognitive architecture is "massively modular." And why their attachment to this thesis often drives them to adaptationism in their speculations about the phylogenesis of cognition. Then we'll have the whole picture in view, and I can tell you what I think is wrong with it. In case you care.

That, however, will come later. I want to spend the rest of this chapter reflecting a little on the notion syntactic structure itself. As we've been seeing, the idea that mental representations have syntactic properties is at the heart of the nexus between rationalist psychology and the computational theory of mind. What, then, are syntactic properties?

What, Then, Are Syntactic Properties?

Well, to begin with: Syntactic properties are peculiar. On the one hand, they're among the "local" properties of representations, which is to say that they are constituted entirely by what parts a representation has and how these parts are arranged. You don't, as it were, have to look "outside" a sentence to see what its syntactic structure is, any more than you have to look outside a word to see how it is spelled. But though it's true that the syntax of a representation is a local property in that sense, it's also true that the syntax of a representation determines certain of its relations to other representations. Syntax, as it were, faces inward and outward at the same time. I want to emphasize this duality since, as we'll see in the chapter 2, both the cardinal virtues and the regrettable limitations of Turing's kind of computational psychology very largely turn on it. For the present expository purposes, I propose to talk about the syntax of sentences rather than the syntax of mental representations; but the morals apply mutatis mutandis assuming that RTM is true.

The grammatical fact that "swims" is the main verb and "John" is its subject in the sentence "John swims" is constituted entirely by facts about what the parts of that sentence are and how they are put together. But this local property of "John swims" nevertheless determines various of its relations to other English sentences: for example, that "who swims" and "does

John swim" are among the question forms of "John swims," but that *"who does John swim" is not. In consequence, if a mechanism were sensitive to the *local* syntactic structure of "John swims," it would thereby be in a position to predict such *relational* properties of the sentence as its having the question forms that it does.

Likewise for the logical form of a sentence (its logical *syntax,* as logical form is sometimes called). That a sentence has the logical form *Fa* is entirely a matter of the identity and arrangement of its parts; but its being of that form nevertheless constrains various of its intersentential relations. For example, if such a sentence is true, so too is the corresponding sentence of the form $\exists x(Fx)$. In consequence, a mechanism that is directly sensitive to the logical form of a sentence is thereby indirectly sensitized to certain of its entailments. It's yet another way of putting Turing's insight that local structure can encode not only grammatical relations among sentences, but inferential relations as well.[18]

Syntactic properties aren't, of course, the only ones that exhibit the kind of internal/external duality just remarked on. Here's a sort of simile, for those of you who may like such things.

Consider the famous ethology of the three-spined stickleback. All we need of it, for present purposes, is that when a male of the species is sexually active, it develops a characteristic red spot (on, approximately, its tummy) to which other sexually active male sticklebacks react with characteristic displays of territorial aggression. Now, being sexually active is a complex, largely dispositional property, the possession of which affects all sorts of relations between a stickleback and its peers. By contrast, having (or not having) a red spot on its tummy is a "local" property of sticklebacks in much the same sense that containing the word "John" is a local property of the sentence "John swims." That a stickleback has a red spot on its tummy is constituted entirely by the identity and arrangement of its parts. Here, then, is the point I want to emphasize: in consequence of the reliability of the relation between being, on the one hand, a sexually active male stickleback and, on the other hand, being a male stickleback with a red patch on its tummy, a mechanism that is able to respond

(directly) to the red patch is *thereby* able to respond (indirectly) to the pattern of behavioral dispositions characteristic of a sexually active male.[19] Uncoincidentally, other male sticklebacks are notable among such mechanisms.

To be sure, this analogy between a sentence's syntax and a stickle back's tummy is imperfect. I want to stress one of the differences because it will turn out to be crucial in later chapters: Whereas the identity and arrangement of its parts is among the *essential* properties of a representation, the color of its tummy is *not* among the essential properties of a stickleback. The identity of a fish generally survives alteration of the color of its tummy, *but the identity of a sentence never survives alterations of its syntax or its logical form.* Thus, a sentence that doesn't contain "John" ipso facto can't be a token of the same type as "John is bald." Likewise a sentence that doesn't entail that someone is bald.

I think perhaps that's enough of chapter 1. We now have in place a continuation of rationalist epistemology that emphasizes inferences from poverty of the stimulus to conclusions about what cognitive contents are innate. And we have a continuation of rationalist psychology that reconstructs both the notion that mental states can have logical forms and the notion that their logical forms can be determinants of their causal powers. It does so by assuming that mental representations have syntactic structures, that the logical form of a thought supervenes on the syntactic form of the corresponding mental representation, and that mental processes are computational in a proprietary sense of "computation" that turns on the notion of a syntactically driven causal relation. So be it.