NOTICE
WARNING CONCERNING
COPYRIGHT RESTRICTIONS

The copyright law of the United States [Title 17, United States Code] governs the making of photocopies or other reproductions of copyrighted material

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that use may be liable for copyright infringement.

This institution reserves the right to refuse to accept a copying order if, in its judgement, fulfillment of the order would involve violation of copyright law. No further reproduction and distribution of this copy is permitted by transmission or any other means.
CHAPTER SIX

On the Impossibility of Acquiring "More Powerful" Structures

The extreme but very consequential stand which is outlined here by Jerry Fodor and which is expounded in due detail in his essay, The Language of Thought, is a stumbling block to all those who, like Papert, Inhelder, Cellérier, and Piaget himself, are in favor of a "compromise" between innatism and constructivism. The grounds for such a compromise have been tentatively, though recurrently, identified in the domain of cognitive "primitives," that is, those early and still poorly differentiated processes out of which specific language structures and specific cognitive strategies might develop. Granted that some structures are innate, it is not on the side of complex, highly specific linguistic rules such as the specified subject condition (SSC) or bound anaphora that one has to look. These rules may be the terminal, sophisticated outcome of a vast class of interactions among simpler primitives, that is, a repertoire of general-purpose computational abilities to be found at the common root of a variety of linguistic and extralinguistic competences. Such a conclusion, which has been offered repeatedly by the Geneva group and by Seymour Papert, forms in addition the building block of Putnam's contribution to this debate. When Piaget insists on the demarcation between properties of the cognitive system which are "compatible with the genes, but not determined by them," when Cellérier confines the genetic component to a "general-purpose self-programming problem solver" capable of creating genuine novelty (see Chapter 2), when Papert claims that all that the genetic specification can allow for is computational
"primitives," the core issue is again and again one of conciliating specificity (of invariant structures) with flexibility (of the overall performance range).

Admitting that a finite set of multipurpose precursor operations (and nothing else) is innate, the stepwise differentiation and integration of such basic structures by the action of the external environment can account for the variety of cognitive and linguistic abilities, making of innatism and constructivism two faces of the same coin. This, in its simplest terms, is the "compromise" between Chomsky and Piaget propounded by many authors. Fodor's argument frustrates these conciliatory hopes at a very fundamental level, pushing the innatist approach into an almost paradoxical position. Nothing new, he contends, can be "acquired" during cognitive development, and this for purely logical reasons. The growth of language, apart from obvious accretion of vocabulary, and the growth of knowledge, apart from obvious accretion of information, are to be seen as the unfolding of predetermined developmental stages involving specialization and restriction of competence. The compromise already rejected by Chomsky on grounds of weakness of the proposed constructivist hypotheses (see Chapter 6) is rejected here by Fodor on more general grounds, because of the inherent fallacy of the concept of "acquiring a more powerful structure." The implications of Fodor's argument concerning the philosophy of mind will appear in sharper focus in his "Reply to Putnam" in Part II.

Fixation of Belief and Concept Acquisition

Jerry Fodor

It seems to me that there is a sense in which there isn't any theory of learning, and this is quite compatible with Chomsky's point that maybe there is no general learning mechanism that applies equally to perception, language, and so on. I'll argue not only that there is no learning theory but that in certain senses there certainly couldn't be; the very idea of concept learning is, I think, confused.
That is a contention that goes against the dominant thought of the last 300 years or so, in both psychology and philosophy. The sort of picture that one is invited to have in philosophy and psychology is that there exist two rough alternatives: there is the nativist alternative, which is methodologically very obscure and maybe unscientific and unintelligible, but in any case, very low down on the hierarchy of possible theories; then there is the learning alternative. The latter is at least supposed to be clear—that is, we know what it would be like to have a theory of learning; we know what it would be like to explain something in terms of learning. The only question is, with what generality can we do so?

Yet it seems to me that nobody has ever presented anything that would have even the basic characteristics of a theory of learning in the sense of a theory of concept acquisition. The reason people think that there is such a theory is that they confuse a theory of concept acquisition with a theory that has a quite different kind of logical structure, what I will call a theory of "fixation of belief." Although there are no theories of concept learning, there are theories of fixation of belief, and I think they are very deep and very important. The trouble is that to get any of them to work you have to be radically innatist about the origin of concepts. In short, no theory of learning that anybody has ever developed is, as far as I can see, a theory that tells you how concepts are acquired; rather, such theories tell you how beliefs are fixed by experiences—they are essentially inductive logics. That kind of mechanism, which shows you how beliefs are fixed by experiences, makes sense only against the background of radical nativism. That is the position that I am going to try to argue for.

For reasons of space I can't set out my point of view in great detail, and I certainly can't make a general case; what I will do is give a sketch of the kind of argument that I have in mind and then give two applications: one to what is known as concept learning experiments in psychology (I have in mind the kind of things that Vygotsky, Bruner, and others have done), and the other to what I consider to be a way of reading the Piagetian system.

Let me say, in the first place, something that I think everybody has always assumed is true in one form or another about theories of learning: I take it that anybody who has ever given a theory of learning in terms of mental processes anybody who has ever said anything about what the information flow in
learning is like] has said, in effect, that learning is a matter of inductive extrapolation, that is, of some form of nondemonstrative inference. It follows that a learning theory is some function that takes you from some sets of beliefs about your experiences into some sort of general beliefs, and secondly, that any such theory must acknowledge, among the processes involved in learning, hypothesis formation and confirmation.

Let's see how this works. Consider the so-called concept learning experiment. The subject enters an experimental situation and the experimenter puts a stack of cards in front of him: there are, let's say, red circles, green triangles, blue tetrahedrons, and so on. The colors and shapes of the objects on the cards are allowed to vary freely. What the subject has to do is to sort the cards into piles which are called "plus" and "minus" (which might also be called "satisfy the experimenter" and "don't satisfy the experimenter," or "reinforced" and "unreinforced," and so forth). Every time the subject [S] sorts the cards, the experimenter [E] tells him whether he is right (for example, he says "You're right," or he nods his head, or he gives the subject an M&M). Eventually, S comes to sort "correctly," that is, to place in a given pile all the cards that E wanted there, and only those cards.

This kind of experiment is sometimes thought of as involving language learning as well as concept learning. You can turn it into a language learning experiment in the following way: instead of being asked to put the card into one or the other pile, the subject is told to say a nonsense word like miv if he thinks the card is of one kind and to say non-miv if he thinks it is of the other kind. You can then think of this either as an experiment in which you learn the word miv, or as an experiment in which you learn the criterial attributes that define the concept miv.

Now, let's consider the theories of what happens in the experiment. There is only one theory that I know of, and it goes like this: the organism (child, adult, rat, or whatever) comes to develop some such hypothesis as "X is miv, if and only if X is . . . ," and then he fills in the blank on the right-hand side with a specification of the attributes that are criterial for being miv, say, red and square. All the experiment tells us is how [for example] various choices of the attributes, or various choices of the reward, or various relations between the two, affect the subject's convergence on accepting the right hypothesis. What it doesn't tell you is where the hypotheses (and the concepts that
they deploy) come from! If you think about the classic concept learning experiments, I think you will see that that is true of every one of them.

To summarize, everybody agrees that S eventually ends up sorting into mivs and non-mivs correctly, and that he does so as a consequence of his history of being told which particular cards are miv and which aren't. There exist, moreover, prototheories of this process; theories that tell us, in effect, with what probability a given body of experience with the cards will lead to the fixation of one or another belief about the extension of miv. Such theories have, as I remarked previously, the general character of inductive logics: that is, they map pairs consisting of hypotheses ("X is miv if X is . . .") and sets of data-statements ("Card number 17, which I was reinforced for calling miv, had a red square on it") onto a number which can be interpreted as the degree to which that body of experiences tends to fix a belief in the truth of that hypothesis. Our psychology of learning details this mapping: it tells how the degree of fixation of belief varies with variations of experimental parameters. I claim that all theories of learning do this, and that none do anything else (that is, anything else relevant to the present discussion).

The difficulty with such a theory is, however, that it says nothing about the origin of concepts. In particular, it assumes as "given" the "criterial attributes" which form the hypotheses that are "fixed" in the experimental situation. In consequence, a theory of how our beliefs are determined by our experiences is not a theory of the source of our inductive hypotheses. On the contrary, it presupposes the availability of such hypotheses [and, of course, of the experiential data] and tells us only how the likelihood that one or another hypothesis will be accepted by an organism varies with one or another aspect of the organism's experience of its environment. I am saying that an inductive logic (that is, a theory of the fixation of beliefs; that is, a theory of learning in the only sense in which there are theories of learning) can't tell you how the concept miv is acquired because it presupposes the availability of that concept when it assumes that miv occurs in the confirmed inductive hypothesis. As far as I know, nobody except the nativists has addressed the question of how the concept miv (that is, red and square) is acquired, and what they have said is that it isn't acquired.

I am not, of course, arguing that we should abandon the notion that a learning device is, in essence, an instantiated inductive logic. Rather, the point is that to let such a device do what it is supposed to do, you have to presuppose the field of
hypothesis, the field of concepts on which the inductive logic operates. In other words, to let this theory do what it is supposed to do you have to be in effect a nativist. You have to be nativistic about the conceptual resources of the organism because the inductive theory of learning simply doesn’t tell you anything about that—it presupposes it—and the inductive theory of learning is the only one we’ve got.

Now I want to say something about how this whole way of looking at things applies to the Piagetian theory, at least as it is interpreted in certain places in the States; say, in my office. It seems to me that the following is at least one way of formulating the Piagetian view: Suppose you are Kant and are interested in writing the “First Critique,” that is, you are interested in characterizing the computational capacities of the organism in terms of some very general constraints on the character of the concepts available to it. One way of reading the Piagetian position is to say that if you did that for several different time slices of the organism (instead of just considering the adult), what you would get is a fundamentally different galaxy of constraints on the organism’s concepts. Moreover, this difference would have the following important characteristic: the logic instantiated by the system of concepts at any $i^{th}$ stage is weaker than the logic instantiated by the $i - 1^{th}$ stage (I take this to be implied by such remarks as children at certain stages “don’t have” transitive concepts, or “don’t have” reversibility, and so on). In short, if you look at the organism as a succession of logics, then it is a succession of increasingly powerful logics, and powerful in some fairly rigorous sense: as for example that the set of truths that could be expressed by using the concepts available at $i$ is a subset of the set of truths that could be expressed by using the concepts available at $i + 1$.

Let’s suppose for the moment that in fact the organism does decompose in exactly that way; that a developing child is a series of logics such that each logic literally contains the preceding one, and “contains” is an asymmetrical relation. The logics, then, get stronger; that is, each successive one has the former one as a proper part. Let’s suppose that is true. Now, what I want to argue is that if it is true, then Piaget must not, in point of logical necessity (not empirical necessity) be a non-nativist about changes of stage, that is, about the mechanisms that take you from one stage to the other. In particular, I want to argue that according to this viewpoint, the change of stage cannot be a learning process. It is an argument fundamentally analogous to the one I made when I discussed concept learning.
Why can't it be a learning process? Well, let's assume, once again, that learning is a matter of inductive inference, that is, a process of hypothesis formation and confirmation. Then the least you would have to be able to do at stage 1 is to characterize truth conditions on formulas containing the concepts in stage 2. You have to be able to do that because in stage 1 you have to have hypotheses about when these concepts are instantiated; as I said, hypothesis formation and confirmation is the only model of learning we've got.

The point is that we have a succession of stages of increasing power and we are going to have learning, that is, hypothesis formation and confirmation, mediating the relation between the stages. Now, how is that going to work? It is immediately obvious, from the following example, that it can't work.

Suppose we have a hypothetical organism for which, at the first stage, the form of logic instantiated is propositional logic. Suppose that at stage 2 the form of logic instantiated is first-order quantificational logic. The particular example does not matter in any respect, except that I want it to be clearly a case of a weaker system at stage 1 followed by a stronger system at stage 2. And, of course, every theorem of a propositional logic is a theorem of first-order quantificational logic, but not vice versa.

Now we are going to try to get from stage 1 to stage 2 by a process of learning, that is, by a process of hypothesis formation and confirmation. Patently, it can't be done. Why? Because to learn quantificational logic we are going to have to learn the truth conditions on such expressions as "(X) Fx." And, to learn those truth conditions, we are going to have to formulate, with the conceptual apparatus available at stage 1, some such hypotheses as "(X) Fx" is true if and only if . . . But of course, such a hypothesis can't be formulated with the conceptual apparatus available at stage 1, that is precisely the respect in which propositional logic is weaker than quantificational logic. Since there isn't any way of giving truth conditions on formulas such as all "(X) Fx" in propositional logic, all you can do is say: they include Fx and Fx and Fx, and so on.

If you think about this, you will see that this is an entirely general form of argument, one that shows that it is never possible to learn a richer logic on the basis of a weaker logic, if what you mean by learning is hypothesis formation and confirmation. Yet I say again that learning must be nondemonstrative inference, there is nothing else for it to be. And the only model of a nondemonstrative inference that has ever been pro-
impossibility of acquiring "more powerful" structures / 149

posed anywhere by anyone is hypothesis formation and confirmation.
I take it from this that there literally isn't such a thing as
the notion of learning a conceptual system richer than the one
that one already has; we simply have no idea of what it would
be like to get from a conceptually impoverished to a conceptu-
ally richer system by anything like a process of learning. Thus
there is an important sense in which the nativist hypothesis is
the only one in the field, and the situation has been put exactly
backwards over the last 300 years. The only intelligible theory
of enrichment of conceptual resources is that it is a function
of maturation, and there simply isn't any theory of how learning
can affect concepts.

Discussion

Papert  Winston's model of learning¹ has the following form:
make a hypothesis and from that compute a modification of
the hypothesis. So it is not a matter of selection or confirma-
tion; it is a process of successively modifying it. I don't know
if I believe Winston's theory of learning, but it certainly
doesn't fit your categorization of what a theory of learning
has to be.

Fodor  It doesn't make any difference that I can see. You say
that in the process of going from here to there, you can have
feedback loops, but that doesn't change the logical situation.
The feedback, after all, must be directed toward rejecting the
hypothesis if it is false or accepting it if it is true; in both
cases the hypothesis has to be available to the learning sys-
tem until it is modified in terms of a better hypothesis, and
then that hypothesis has to be available. It just confuses the
issue to ask whether or not there are feedback mechanisms
involved. Assuming that the mechanisms are a kind of induct-
ive logic, any kind of inductive logic is going to have this
problem.

Piaget  I was very much interested by Fodor's presentation, and
I fully agree with Fodor on the first part, that is, on the im-
possibility of explaining language development by means of
a theory of learning in the usual sense of the term; but the
way he reduces my ideas to a theory of learning appears to me