Chapter 10

The Present Status of the Innateness Controversy

Plato says in Phaedo that our 'imaginary ideas' arise from the preexistence of the soul, are not derivable from experience—read monkeys for preexistence.

--Charles Darwin

Introduction

Disclaimers first. Many of you may have been misled by the title of this paper to suppose that it is about the current status of the innateness controversy. Well, barely. In fact, my topic is restricted to questions about the innateness—or otherwise—of concepts. I shall therefore have nothing whatever to say about the issue of the innate determination—or otherwise—of differences in mental traits and capacities. (I assume that the psychology of individual differences is largely vacuous for our species, as it patently is for every other.) I shall also have nothing to say about the current controversy over the innateness of beliefs, except insofar as the holding of certain beliefs may be thought to be constitutive of the attainment of certain concepts. (For example, it might be held that you can’t, in point of logic, have the concept BACHELOR unless you

I have borrowed several of the ideas in this paper—and a few of the jokes—from Fodor, Garrett, Walker and Parkes', "Against definitions" (1980). My indebtedness to my co-authors is hereby gratefully acknowledged. Much of this material was first presented as the 1979 Walker-Ames Lectures at the University of Washington. I am most grateful to friends and colleagues there—and especially to Professor Charles Marks—for their stimulating comments and discussion.
have the belief that bachelors are unmarried. If this is true, then the question whether the concept BACHELOR is innate willy-nilly involves the question whether the belief that bachelors are unmarried is.) It does seem to me that Chomsky’s demonstration that there is serious evidence for the innateness of what he calls ‘General Linguistic Theory’ is the existence proof for the possibility of a cognitive science; indeed, that it is quite possibly the only important result in the field to date. But I shan’t discuss either Chomsky’s claim or his evidence beyond remarking that if he’s right, then the innateness of some concepts—viz. the ones that figure in the innate beliefs—follows trivially. You can’t have innately the belief that, say, transformations must apply in cycles (or, for that matter, the belief that they must not) unless you have innately the concepts TRANSFORMATION and CYCLE.

I should also confess that, though some of the following may sound like philosophy—especially to nonphilosophers—it is really only a kind of speculative psychology. The issues I’ll deal with are empirical issues (or so I’ll maintain) and quite a lot of the apparatus of current cognitive psychology will be taken for granted; without explication even. I’ll assume, for example, that the mentalistic notion of a concept can be made clear enough to be scientifically respectable, since, if it can’t, the question whether any concepts are innate won’t be worth discussing. I assume this, moreover, in the teeth of prevailing philosophical opinion to the contrary. My excuse is that I can’t imagine a serious psychology which dispenses with that notion, and neither, apparently, can the psychologists.

So much for the disclaimers. What I propose to do in this paper is to consider the innateness issue as it arises in the context of a certain broad class of theories of concept attainment which I will call classical theories of concept attainment. Not all conceivable theories of concept attainment are classical theories, and the assumptions that classical theories share are philosophically—and otherwise—tendentious. Still, in point of historical fact, just about every theory of concept attainment that any psychologist or philosopher has succeeded in taking seriously, barring only behavioristic theories, counts as classical by my criteria. So it would be interesting to show—what I propose to maintain—that the most plausible classical theories are Nativistic. At a minimum, I’ll try to convince you that the issue between Nativist and Empiricist versions of classical theory has been widely misrepresented and that when one sees what the issue really is, one sees that the Nativist account is distinctly in the running.

I need, by way of stage setting, to say something about what makes a theory of concept attainment a classical theory of concept attainment. Since I don’t propose to defend the classical theory, and since I want to fish with a big net, the account I’ll give will be rough and sketchy.

To begin with, classical theories are ipso facto species of the Representational Theory of Mind (RTM). The Representational Theory of Mind holds either that propositional attitudes are relations between organisms and mental representations or that such relations are nomologically necessary for having a propositional attitude. (That is, RTM may be, but needn’t be, held as an ontological doctrine.) Mental representations are mental objects: RTM quantifies over them like mad. Typical claims that RTM holds about mental representations are these: there is an infinite proper subset of which are internally structured; mental representations exhibit such semantic properties as truth and reference; mental representations constitute the domain of mental processes.

So, for example, a well known version of RTM had it that mental representations are mental images; that mental images refer to what they resemble; that to believe that such and such is to bear a certain kind of relation to a mental image of such and such; that some mental images have other mental images as proper parts; and that associative processes apply to mental images in virtue of relations of co-occurrence, resemblance, etc., that hold among the images. I suppose that Hume held RTM in something like this form; it seems certain that some of Clark Hull’s followers did.

A second thing you have to accept to hold a classical theory of concept acquisition is that concepts are mental entities; in particular, that concepts are species of mental representations. Roughly, a concept is that sort of mental representation which expresses a property and is expressed by an open sentence. So, for example, if you hold the classical theory (hence RTM) on the image version, you will think of the concept DOG as (a) an image of a dog; (b) expressing the property of dogness (and hence applying to dogs); and (c) as being the mental representation...
normally evoked by utterances of the form of words ‘... is a dog.’ That’s as precise as I propose to get, this being Cognitive Science. It is worth emphasizing, however, that the mental entity account of concepts is independent of, not incompatible with, Frege’s view that a concept is an abstract entity: a Sense. Qua mental representations, concepts have semantic properties. It is thus open to you, if you’re so inclined, to run RTM (hence classical theory) on a Fregean line. To wit: the word ‘dog’ expresses the concept DOG; the concept DOG expresses the sense dog (i.e., expresses a certain Fregean concept); the relation expresses is transitive. On this account (to which I am, in fact, partial) expressing a Fregean concept is a semantic property par excellence, and the formulae of natural language inherit their semantic properties from those of mental representations. From now on, when I speak of attaining a concept I shall have the mental-object sense of “concept” in mind. It is an agreeable suggestion that attaining a concept (mental-object sense) is de facto a necessary condition of becoming acquainted with a concept (Fregean sense); i.e., that our acquaintance with Fregean concepts is mediated by our mental representations. But classical theories of concept attainment are not, per se, required to endorse this suggestion.

All classical theories of concept attainment are species of the RTM, and they all take concepts to be species of mental representations. Classical theories of concept attainment come in two varieties: the Empiricist and the Nativist. So far I have been concerned with what these two varieties of classical theories of concept attainment have in common. In the following Section, we’ll start to see what distinguishes between them.

Where Do Concepts Come from?

A concept is what an open sentence of a natural language expresses. If we fix the language (choosing, say, English), we can distinguish between lexical concepts and phrasal concepts. By stipulation, a lexical concept is one that is expressible, in English, by an open sentence with a morphologically simple—viz. a monomorphic—predicate term. (So, for example, GREEN is a lexical concept, since it is expressible by the open sentence ‘... is green’. I shall stretch a point in favor of predicate nominals and consider BACHELOR to be a lexical concept because it is expressible by the open sentence ‘... is a bachelor.’) Whereas, a phrasal concept is one that is expressible, in English, only by an open sentence with a morphologically complex predicate term. (So, for example, the concept LIVES IN CHICAGO AND EATS MANGLEWORTS is a phrasal concept, since, to put it roughly, there is no one-word predicate term of English which means lives in Chicago and eats mangleworts, and IS UNCOMMUNICATIVE is phrasal because ‘uncommunicative’ is multimorphemic).

For a number of reasons, the lexical/phrasal distinction won’t bear much theoretical weight. For example, we would presumably get a different partitioning of the concepts if we changed the reference language; for all I know, some concepts that are expressible only by phrases in English are expressible monomorphically in, e.g., Urdu. And, for all I know, some concepts aren’t expressible in English at all, so perhaps the lexical/phrasal distinction isn’t exhaustive. Never mind; I only want the distinction for heuristic purposes. We’ll see how to dispense with it as the discussion goes along.

We arrive at the first consideration of substance: all theories of concept attainment assume that there are infinitely many phrasal concepts which are acquired by the application of constructive procedures to a basis of lexical concepts. Among the governing considerations are these:

(1) It seems clear that the semantic properties of phrasal concepts are typically inherited from the semantic properties of the corresponding lexical concepts. So, patently, the concept expressed by, say, ‘the cat on the mat’ inherits its semantic properties from (inter alia) those of the concept expressed by ‘cat’. It’s thus plausible to think of the latter concept as literally a constituent of the former (conversely, to think of the former concept as literally a construct out of the latter) thereby preparing the groundwork for explaining why nothing satisfies THE CAT ON THE MAT unless it satisfies CAT.

(2) Not only do the semantic properties of phrasal concepts typically depend on the semantic properties of corresponding lexical ones, but the dependence appears to be determined by general and regular principles: on the one hand, THE CAT ON THE MAT differs from THE DOG ON THE MAT in much the same way as THE CAT differs from THE DOG; and, on the other hand, the semantic properties of THE CAT ON THE MAT seem to depend on the semantic properties of THE CAT in much the
same way that the semantic properties of THE DOG IN THE MANGER depend on the semantic properties of THE DOG. All of which suggests that the principles that construct phrasal concepts from lexical concepts are fairly general in application—in particular, that they can be formulated without listing the concepts they apply to.

(3) A conceptualist theory of language comprehension will presumably want to say that understanding an utterance involves entertaining the concept that it expresses. It comports with that sort of treatment to suppose that the concept expressed by a phrase is a construct out of the concepts expressed by the lexical constituents of that phrase. Attainment of the appropriate lexical concepts, and of the principles of construction, will then constitute a sufficient basis for attaining the corresponding phrasal concepts. Indeed, it’s natural to suppose that the “logical syntax” of a phrase exhibits the internal structure of the concept it expresses: that it exhibits the relation of the phrasal concept to the lexical concepts it’s constructed from.

This is all familiar stuff and, so far as I know, it’s uncontroversial given the framework of RTM. I’ll assume, therefore, that all theories of concept attainment tell more or less the same story about phrasal concepts: there is a (finite) basis of lexical concepts and there is a constructive apparatus. Phrasal concepts are constructed from lexical concepts by application of combinatorial rules. Phrasal concepts are thus reducible to lexical concepts, save their semantic properties. Since the rules of construction are recursive (in the sense of ‘iterative’), the set of phrasal concepts is productive.

The various classical theories of concept attainment differ (inter alia) in their formulations of the constructive principles. Some earlier versions take them to be principles of association; some later ones borrow them from the logical theory of quantification or from the syntax of computer languages. For present purposes, however, such differences needn’t much concern us. What does bear emphasis is that, on any such account, attaining a phrasal concept will involve actually performing the relevant logical construction: to attain the concept BROWN COW will involve actually assembling it from the constituent concepts BROWN and COW. (Someone who has attained the latter concepts but has not performed the construction will have the former concept only potentially). Similarly, thinking about brown cows will involve entertaining a complex mental representation of which the concepts BROWN and COW are constituents in whatever sense of constituency the theory favors.

In short, the phrasal concepts are the closure of the lexical concepts under some combinatorial apparatus. What about the lexical concepts? It is here that interesting differences between Empiricist and Nativist treatments begin to emerge. I want to spell out what I take to be the Empiricist view at some length.

For the Empiricist, the lexical concepts are a mixed lot. Just as phrasal concepts are patently complex relative to their constituent lexical concepts, so, on the Empiricist story, very many lexical concepts are complex relative to a set of mental representations that we can call the primitive conceptual basis. Some of the essential Empiricist claims in this area might be put as follows:

(4) All concepts—lexical and phrasal—are either primitive or constructed from primitive concepts by a combinatorial apparatus. It’s usually assumed that the same combinatorial apparatus that constructs phrasal concepts from lexical ones also operates to construct complex lexical concepts from the primitive ones. It’s thus a typical Empiricist view that the internal structure of complex lexical concepts mirrors the internal structure of phrasal concepts. Indeed, it’s usual to suppose that there are some—possible many—cases in which a word and a phrase express the same concept; “bachelor” and “unmarried man” are both said to express the concept UNMARRIED MAN, and, in general, each definable word is said to express the same concept as the phrase that defines it.

(5) Whereas the distinction between lexical and phrasal concepts is largely accidental—many lexical concepts have internal structure, just as phrasal concepts do—the distinction between primitive concepts and complex concepts is epistemologically principled. That is, we can pick out the primitive concepts by some epistemologically interesting criterion, and it will turn out that those, and only those, concepts which satisfy the criterion are the ones that lack internal structure.

(6) The primitive conceptual basis is much smaller than the set of concepts that are lexicalized by any natural language; for example, it is much smaller than the set of concepts that happen
to be expressed by morphemically simple predicate terms in English.

Empiricists disagree as to precisely how one ought to formulate the criterion that picks out the primitive (hence internally unstructured) concepts. But the rough idea is that primitive concepts are sensory. Let's, for the moment, not worry very much about what that comes to. I'll simply assume that every organism whose mental states are worth discussing has a sensorium—perhaps a set of receptive mechanisms that satisfy the technical notion of a transducer—and that the primitive concepts are the possible outputs of the sensorium across the range of inputs that it responds to. That is, specify the possible outputs of the sensorium and you have the primitive concepts that the organism can potentially entertain. Take the closure of the primitive concepts under the combinatorial apparatus, and you have all the concepts that the organism can potentially entertain.

Figure 1 gives a schematic representation of the Empiricist story about the space of concepts potentially available to an organism. Every concept is either sensory (hence primitive, hence internally unstructured, hence the output of the sensorium for some specifiable input) or constructable from the primitive concepts by application of the combinatorial apparatus. Since this disjunction is exhaustive, the "Empiricist maxim" follows: nothing is in the mind except what is first in the senses. The complex concepts include lexical and phrasal concepts indiscriminately, except that no primitive concept is phrasal. (Remember that the logical syntax of a phrase displays the internal structure of the concept it expresses. A concept with no internal structure cannot, therefore, be phrasal, since every phrase has a logical syntax—excepting only idioms.)

The space of potentially available concepts is fixed given an enumeration of the sensory concepts and a specification of the combinatorial mechanisms. I might equally have said: the space of potentially available concepts is fixed given a specification of the combinatorial mechanisms and a characterization of the sensorium. From our point of view, a sensorium is simply the realization of a function from stimuli onto primitive concepts. Since the Empiricist assumes that primitive concepts can arise only via activation of the sensorium, we can think of a theory of the sensorium as exhaustively specifying the causal occasions for the availability of any primitive concept: the structure of the sensorium determines what sort of stimulation is causally necessary and sufficient for the attainment of each such concept. It's going to be an important difference between Empiricist and Nativist accounts that, according to the latter, primitive (viz. internally unstructured) concepts may become available by means other than the activation of the sensorium.

It's important to understand that Figure 1 represents the Empiricist doctrine about what concepts are potentially available to an organism, not the Empiricist doctrine about how concepts are attained. Figure 1 says: Here's the class of stimuli which can, in point of nomological fact, affect the conceptual repertoire of the organism; and here's the class of concepts which the organism can, in point of nomological fact, entertain in consequence of such stimulations. Whereas a theory of concept attainment must explain how the concepts that an organism actually acquires are determined by the stimuli that it actually encounters: it specifies the mechanisms in virtue of which the conceptual repertoire of an organism varies as a function of its experience. It's this latter sort of theory that we are primarily concerned with in what follows.

Empiricist theories of concept attainment distinguish sharply between the mechanisms postulated to explain the acquisition of primitive concepts and those postulated for the acquisition of complex concepts (be they lexical or phrasal). We've seen that the account of the former mechanisms is, in effect, embodied in the theory of the sensorium. What about the acquisition of complex concepts?

It's here that the concept-attainment models familiar from the psychological and AI literature come into play. Perhaps the best
way to understand this literature is to consider what happens in the sorts of experiments that psychologists use to study what they call “concept learning”. (Terminological note: from here on, I shall use 'concept learning' as a technical term for the acquisition of a concept as a result of the particular mental processes now to be described. I shall use 'concept acquisition'—or 'concept attainment'—for any mental processes which eventuate in the availability of a concept to an organism. The terminology thus means to leave it open that some (many, all) concepts might be acquired but not learned).

A typical concept learning experiment goes as follows (for purposes of exposition, you get to be the Subject): You arrive in the experimental environment. The experimenter says to you something like, “I have here a pack of stimulus cards. Each stimulus card displays a colored geometrical figure. There are some red triangles, there are some green squares, there are some yellow circles, there are some red rectangles, and so on. Now, some of these cards are flurg and some of them are non-flurg and (for convenience) every card is either flurg or non-flurg. Your job is to figure out which cards are which. And I am here to help you. For, though I will not tell you which ones the flurg cards are, still I will let you examine any card you like and guess whether it is flurg. And I will tell you if your guess is right. Proceed”.

You proceed. That is, you examine a card (say it is a green triangular one) and you guess (say you guess flurg). Now, unknownst to you, the experimenter has in mind a criterion for correct responding; for example, that the response “flurg” is correct iff the card is either green or square. Since the card you drew was either green or square (it was green and triangular), and since you guessed flurg, the experimenter says “Right you are” and you are allowed to go on to the next card. You continue in this wise until, mirabile dictu, your performance “reaches criterion”; e.g. until you say “flurg” for all green or square cards and “non-flurg” for all the rest on, as it might be, twenty consecutive trials. You are then said to have learned (a fortiori attained) the concept flurg and you are allowed to go home.

Various aspects of this scenario are inessential to its being a concept-learning experiment. For example, the stimuli didn’t have to be colored shapes; they could, in principle, have been any discriminable array. It’s also inessential that the criterial property be a Boolean function of the stimulus features; the experimenter could have said “right” depending upon the ordinal position of the stimulus card in the series. It’s also inessential that the experimenter uses a verbal signal to tell you which of your responses are correct; he could equally well nod and smile, or supply you with an M & M, or shock you when you get it wrong. Indeed, so far as the basic paradigm is concerned, it’s inessential that you’re a person. You might have been a pigeon, in which case we would have called it a “discrimination-learning” experiment and we’d have replaced the verbal instructions by some procedure which “shapes” a response (say, pecking the chosen card) that you could then use to indicate which stimulus belongs in which class. The point is that the general structure of this experiment informs practically all the psychological work on concept attainment which the Empiricist tradition has inspired. Much of what we now know about concept learning has consisted in determining how the organism’s ability to achieve criterion varies as a function of variation in (e.g.) the character of the stimuli, the character of the reinforcement (feedback signal), the scheduling of the reinforcement, the species, age, sex, IQ, traumatic impairment, etc., of the organism, and so forth.

That, then, is the concept-learning experiment. I assume that the artifactual features are eliminable: that the experimental situation does, indeed, engage the mechanisms of concept learning insofar as the concept is complex and insofar as it is learned “from experience”. What, then, is available by way of a theory of the organism’s performance in the concept-learning experiment?

So far as I know, only one such theory has ever been seriously proposed (though a variety of rather different vocabularies have been used to express it). The theory is this: concept learning involves the inductive fixation of certain beliefs, and the mechanisms of concept learning are realizations of some species of inductive logic. In particular, they involve the formulation and confirmation of hypotheses about the identity of the concept being learned.

So, for example, what goes on in your head in the experimental situation we’ve imagined might be something like this: You make your first guess—the green and triangular card is flurg—at random. (For all you know it might just as well be non-flurg, but you have
to start somewhere, so why not flip a coin?) Since, as it turned out, that guess was right, you have evidence for any of a range of hypotheses: e.g. that the flurg cards are the green ones; that the flurg cards are the triangular ones; that the flurg cards are the green and triangular ones; that the flurg cards are the green or square ones; that the flurg cards are the non-red ones ... etc. You pick one (or, perhaps a set) of these hypotheses and you try it. For example, you pick the hypothesis that the flurg cards are the green ones. Then, if the next card is a red square, you say “non-flurg” and the experimenter says “wrong”. The range of hypotheses compatible with your data (i.e. with the outcomes of past trials) is now altered: for example, the hypothesis flurg = green is out, though the hypothesis flurg = red or triangle is still in. So you pick one of the hypotheses compatible with this expanded set of data and you try it; and so on till you find a hypothesis that’s compatible with the old data and which further trials decline to disconfirm. You carry on with this new hypothesis until criterion is achieved.

In short, each trial provides inductive evidence pro or con a hypothesis of the form: the concept being learned is the concept of something which is ..., where what goes in the blank is a candidate specification of those properties of a stimulus in virtue of which it satisfies the concept: e.g. circular and orange, (not red) and rectangular), fifth stimulus in the series, or whatever. The mechanisms of concept learning—insofar as concept learning is assumed to occur in the concept-learning experiment—are thus seen to be the mechanisms of inductive inference.

About which a fair bit is known. A mechanism for performing inductive inferences is ipso facto a (realized) inductive logic. As such, it must have access to at least the following:

(7) A source of hypotheses to be tested; which in turn presupposes a canonical format and vocabulary for the specification of the hypotheses.

(8) A canonical format and vocabulary for the representation of the data (e.g. on trial n the card was circular and the response “non-flurg” was rewarded.)

(9) An a priori ranking of the hypotheses in order of relative simplicity. (Roughly because, barring such a ranking, a body of data bestows equal confirmation on all of the infinitely many nonequivalent hypotheses with which it is formally compatible.)

(10) A confirmation metric: roughly, a function from pairs consisting of hypotheses and data sets onto numbers that express the level of confirmation of the hypothesis relative to the data.

Let me emphasize that 7–10 are presupposed by an induction machine; they represent minimal structural commitments for a device that can learn by the projection and confirmation of hypotheses. It is, for example, because bricks do not have functional components corresponding to 7–10, whereas pigeons presumably do, that pigeons but not bricks can learn things.

The present claims are, then, as follows:

(11) All Empiricist theories hold that the concept-learning experiment typifies situations in which (complex) concepts are learned from experience.

(12) All Empiricist theories view performance in the concept-learning experiment as mediated by hypothesis formation and confirmation.

(13) All Empiricist theories are therefore committed to the view that the mechanisms of concept learning constitute a realized inductive logic. (Take-home exercise: Convince yourself that 11–13 are true of, inter alia, “habit-strength” models of associative learning.)

At this point, you may be feeling tempted to argue like this: “Look, if that’s the Empiricist’s model of concept acquisition, then the Empiricist has no model of concept acquisition. For consider: on this view, learning the concept FLURG is learning that a certain hypothesis is true; viz. the hypothesis that the concept FLURG is the concept of something that is green or square. But learning that that hypothesis is true is in turn said to be a matter of entertaining that hypothesis and comparing it with the data. But that hypothesis itself contains the concept GREEN OR SQUARE. Now, surely, if a concept is available to you for hypothesis formation, then you have that concept. And if you have to have a concept in order to perform the concept-learning task, then what goes on in the concept-learning task cannot be the learning of that concept. So the concept-learning task is not a task in which concepts are learned, and the theory of the concept-learning task is not a theory of concept attainment”.

This is a temptation to which, I think, one ought to succumb.
I did so in The Language of Thought, where I argued that if there is an Empiricist theory of concept acquisition, it must consist of an account of the origin of the hypotheses which (by assumption) get inductively confirmed in the concept-learning task. And, de facto, there exists no such theory. (This is hardly surprising; standard accounts of inductive inference presuppose a range of hypotheses to be evaluated, the source of these hypotheses being unspecified. That is one reason why inductive logics are viewed as theories of confirmation, not theories of scientific discovery.)

What has happened is that the Empiricist story recruits what is really a theory of the fixation of belief to do double duty as a theory of the attainment of concepts. This strategy doesn’t work, and the strain shows in all sorts of places. For example, it’s surely clear that any normal adult would have acquired such workaday concepts as GREEN OR SQUARE long before he encountered a concept-learning experiment; hence, achieving criterion in such experiments couldn’t, in the general case, require that any concept be acquired in the course of performing the experimental task. What the subject would have learned in the case described above, for example, is not GREEN OR SQUARE, but only the fact that the experimenter has decided to call things that are green or square “flurg” for the duration of the run—a fact that is interesting only because it controls the distribution of rewards.

I really do think that the Empiricist will have to admit all this: inductive logics are question-begging when proposed as theories of concept attainment, however illuminating they may be as theories of the experiential fixation of belief. Still, having conceded the spirit of his account, the Empiricist might make a reply that would preserve something of its letter. I want to consider this because it raises a point that’s important in the context of the innateness issue. (To which, I promise, we shall presently return. Patience.)

The argument three paragraphs back turned on the assumption that “if a concept is available to you for hypothesis formation, then you have that concept”. And, though there is a sense in which this surely has to be admitted, there is also a sense in which it might be denied. So, it could be held that concept learning does sometimes occur in the concept-learning task, and that when it does, what happens is that a concept previously available to the organism only for the formation of inductive hypotheses then becomes available for a wider range of cognitive tasks (e.g. in reasoning, in the organization of memory, in perceptual identification, etc.). Assume, for example, that there is a sort of ‘faculty of hypothesis formation’ (perhaps it’s a department in the faculty of imagination mentioned in note 3). Concepts like GREEN OR SQUARE must be available to this faculty for purposes of projecting candidate hypotheses in the concept-learning task. Insofar as these concepts are “learned from experience” in the course of performing that task, what happens is that they are made available for general cognitive employment as a causal consequence of the operation of such inductive mechanisms as 7–10 enumerate.

I have no idea whether anything like this is true. I mention the point partly because it saves the Empiricist story from the charge of utter logical vacuity, but also, more importantly, because it serves to emphasize a characteristic difference between the Empiricist treatment of complex concepts on the one hand and of primitive ones on the other. The point to notice is that the Empiricist story about concept learning, even in this revised form, could in principle apply only to the acquisition of complex concepts. This is important because it’s going to be one of my main morals that there isn’t, and couldn’t be, a coherent story according to which primitive concepts are learned.

Learning a complex concept by hypothesis confirmation presupposes the availability of the concept for the expression of the relevant hypothesis. But it’s at least plausible that it needn’t presuppose the availability of that concept for the canonical representation of the data base. Consider FLURG again. The concept FLURG is the concept GREEN OR SQUARE. To learn the concept FLURG is to learn that the hypothesis the concept FLURG is the concept of something that is green or square is true. This hypothesis is confirmed, inter alia, by such observations as stimulus No. 1 is green and is positive; stimulus No. 1 is square and is positive; stimulus No. 7 is red and is negative; etc. The point is that we presumably don’t need, in order to express the data base, such observation statements as: stimulus 1 is green or square and is positive. If we did need such statements, then we would need the concept being “learned” even to express the observations that fix the relevant hypothesis, and the learning model would be entirely vacuous. Another way of putting this is that the constructive
apparatus which generates complex concepts out of primitive ones patently must be available to determine the space of hypotheses that the inductive logic confirms; but perhaps it needn’t also be available to specify the observation base that figures in the inductions. Perhaps, for the latter purposes, the primitive conceptual basis will do.

Now consider the analogous case for learning primitive concepts. Suppose that RED is a primitive concept. Then the hypothesis that has to be confirmed will be: the concept being learned is the concept of something which is RED and will thus, as usual, contain the concept whose acquisition is to be explained. But worse, consider the experiences which fix that hypothesis. They will presumably have to be of the form stimulus No. 1 was red and was positive; stimulus No. 2 was green and was negative, etc. That is, the availability of the concept being acquired will have to be presupposed even to give canonical expression to the experiences that fix that concept. The point is that hypotheses which contain complex concepts can (maybe) be inductively confirmed by a data base in which only primitive concepts figure. But a data base which contains not even primitive concepts is empty and confirms nothing.

Perhaps for this reason, Empiricist models invariably do assume that the concept-learning story applies only to the acquisition of complex concepts, primitive concepts being acquired in some other way. In fact, we have already seen what this “other way” is: whereas complex concepts are acquired by a process of hypothesis formation and confirmation, primitive concepts are assumed to be sensory, and their acquisition is occasioned by the activation of the sensorium.

This difference between the treatment of complex and primitive concepts is highly characteristic of Empiricist theorizing about concept attainment. While the processes which occasion the availability of primitive concepts are, as it were, viewed as merely causal—whether a given pattern of stimulation occasions the availability of a given primitive concept being simply a matter of the structure of the sensorium—the processes which give rise to the acquisition of complex concepts are typically viewed as rational in that the experiences which eventuate in the availability of such a concept are held to bear a confirmation relation to some hypothesis which specifies the internal structure of the concept.

To summarize: the Empiricist model says that there are two sorts of causal processes at play in concept acquisition: in the case of the primitive concepts there are, one might say, brute-causal processes. The structure of the sensorium is such that certain inputs trigger the availability of certain concepts. Punkt. Whereas, in the case of the attainment of complex concepts, there are, as one might say, rational-causal processes: the psychological mechanisms which mediate the availability of such concepts realize an inductive logic, and the experiences of the organism stand in a confirmation relation to the concepts whose availability they occasion.

So much for the Empirist account of concept acquisition. The Nativist account (I shall sometimes say: the Rationalist account, since it received its clearest exposition in the hands of philosophers like Descartes) is very much simpler. To a first approximation, a Nativist says that the mechanism underlying the acquisition of all lexical concepts is brute-causal. Or, to put the same claim slightly otherwise, whereas the Empirist says that many lexical concepts are logical constructs out of primitive concepts which are, in turn, made available by the activation of the sensorium, the Nativist says that the triggering of the sensorium is, normally, causally necessary and sufficient for the availability of all concepts except those that are patently phrasal.

So, for example, an Empirist might give the following ontogeny for the concept TRIANGLE. Primitive concepts of line and angle are available as the causal consequence of sensory stimulation. The concept TRIANGLE is a logical construct out of these primitive concepts. Learning the concept TRIANGLE is a matter of confirming some such hypothesis as the concept TRIANGLE is the concept of something which is ..., where the blank is filled by something tantamount to a definition of “triangle”. Whereas, the Nativist story might go like this: there are trigger stimuli sufficient to occasion the availability of such concepts as LINE and ANGLE; and there are also trigger stimuli sufficient to occasion the availability of such concepts as TRIANGLE. Insofar as one acquires the concept TRIANGLE “from experience”, one normally does so as a consequence of being exposed to triggering stimuli of this latter kind. It may be that the concept TRIANGLE can be viewed
as a logical construct out of the concepts LINE and ANGLE, but even if it is, performing the construction (e.g. confirming a hypothesis that exhibits the structure of the concept) is not part of the learning of the concept.

I'm not entirely happy with this way of putting the issue between Nativist and Empiricist accounts, for reasons that will be made clear later. But suffice it for the moment to notice this: the Empiricist story can't be right unless many concepts that are de facto lexical (concepts like TRIANGLE, BACHELOR, TRUMPET, CIGAR, ELECTRON, and literally tens of thousands of others) are structurally complex. In fact, the unvarnished Empiricist story requires that all de facto lexical concepts except the ones that are sensory (hence primitive) must be structurally complex. Whereas, it's compatible with the Nativist story that all lexical concepts are primitive. This is, I think, close to the heart of the matter, and we'll return to it presently.

First, let's consider in some detail what the Nativist and Empiricist stories do not disagree about; what I want to emphasize is that the point of the controversy has been widely misrepresented and still more widely misunderstood.

Both theories acknowledge a distinction between primitive and complex concepts, and both assume that complex concepts are built out of primitive ones by the operation of some sort of combinatorial apparatus. For example, on both stories, a concept like PEOPLE WHO LIVE IN CHICAGO AND EAT MANGLEWORTS is patently complex, and it inherits its semantic properties from such constituent concepts as PEOPLE, LIVE IN, CHICAGO, MANGLEWORTS, and so forth. For reasons previously considered, both theories will also assume that if a concept is phrasal, it's likely to be complex.

Both theories assume that the availability of primitive concepts is a function of environmental stimulation. In particular, they both assume that the acquisition of such concepts is contingent upon the activation of the sensorium. As Descartes frequently emphasized, it is not part of the Rationalist program to claim that your repertoire of primitive concepts is independent of the character of your experience. On the contrary, just as there are presumably triggering experiences that are nomologically necessary for the acquisition of sensory-primitive concepts like RED, so are there presumably triggering experiences that are nomologically necessary for the acquisition of non-sensory primitive concepts like TRIANGLE. In either case, if you don't encounter the trigger, you don't get the concept.

(16) Both stories assume that primitive concepts are, in a certain sense, unlearned, indeed that they are, in a certain sense, innate.

There are two points to consider here: First, we've seen that the mechanisms of "concept learning" in the technical sense of that notion apply, at best, only to the acquisition of complex concepts. Now there are, in principle, many ways other than learning in which the repertoire of mental representations available to an organism might be affected by its experiences. Think of being hit on your head, or of having your cortex surgically rewired, etc. If you are prepared to suppose—what seems to me quite plausible—that the distinguishing characteristic of concept learning, in the non-technical, pre-theoretic sense of that notion, is the existence of a rational relation between the concept acquired and the experience that occasions its availability, then it will follow on both the Empiricist and the Nativist model that primitive concepts are ipso facto unlearned. According to the Empiricist, complex concepts are learned from experience; but even the Empiricists hold that primitive concepts are merely triggered by experience. And triggering is a non-rational, hence non-learning, mechanism.

If this sounds too much like a terminological quibble, consider the following. On all standard theories, including Empiricist theories, the structure of the sensorium is taken to be innately fixed. It is, for example, just a fact about the way that we are put together that the sensory concepts we have are dependent in the ways that they are upon the particular stimulations which occasion them: if our sensoria were differently constructed, we would have a different repertoire of sensory concepts. So, for example, as Empiricists use the notion "sensory concept", blind men have no sensory concept of red. Nor can they learn that concept; their sensory mechanisms fail to respond to the relevant stimulations. Similarly, in the intended sense of "sensory concept", none of us have sensory concepts of ultrahigh frequency sounds; hence, for none of us is ULTRAHIGH FREQUENCY SOUND a primitive concept. There are, to be sure, imaginable organisms (bats, dolphins,
Martians)—organisms whose sensoria are constructed differently from our own—which do have sensory concepts in this area, and for them ULTRAHIGH FREQUENCY SOUND is primitive for all we know.

This is a kind of point on which classical Empiricists tend to be quite explicit, and none more so, perhaps, than Locke: "... though we cannot believe it impossible to God, to make a Creature with other organs and more ways to convey into the understanding the notice of corporeal things, than those five... which he has given to Man: Yet I think it is not possible, for any one to imagine any other qualities in bodies, howsoever constituted, whereby they can be taken notice of, besides sounds, tastes, smells, visible and tangible qualities. And had Mankind been made with but four senses, the qualities then, which are the object of the fifth sense, had been as far from our notice, imagination and conception, as now any belonging to a sixth, seventh, or eighth sense can possible be: which, whether yet some other creatures, in some other parts of this vast and stupendous (sic) universe, may not have, will be a great presumption to deny" (Locke, 1975, p. 120). Once again, on all standard theories the sensorium is an innately specified function from stimuli onto primitive concepts. The range of potentially available primitive-sensory concepts is thus fixed when you specify the biological endowment of the organism, as is the range of stimulations which can occasion the availability of those concepts.

Attributing to the Empiricists—of all people—the doctrine that primitive concepts are innate may strike you as Quixotic; but I think that is because the nature of the Empiricist/Rationalist controversy has been generally misconstrued in modern readings of the texts. In fact, the idea of the innateness of primitive concepts is widely articulated in the Empirist literature. Here's Hume: "... understanding by innate what is original or copied from no precedent perception, then may we assert that all our impressions are innate, and our ideas not innate" (1909, p. 22). In interpreting this fragment, bear in mind that for Hume simple ideas are mere copies of the corresponding impressions; so, if the latter are innate, so too must be the content of the former.

And William James, whom surely nobody could accuse of wild-eyed Nativism, is still more explicit: "The first thing I have to say is that all schools (however they otherwise differ) must allow that the elementary qualities of cold, heat, pleasure, pain, red, blue, sound, silence, etc., are original, innate or a priori properties of our subjective nature, even though they should require the touch of experience to waken them into actual consciousness, and should slumber to all eternity without it" (1890, Vol. 2, p. 618).

Let me reiterate what I take to be the central point. Primitive concepts are, ipso facto, not learned; and both the range of potentially available primitive concepts and the ways in which their actual availability depends upon triggering stimulation are determined by the innate structure of the sensorium. In these respects, all Empiricists accept the innateness of the primitive conceptual repertoire. What Empiricists don't accept—what, in fact, nobody accepts—is that the availability of primitive concepts is independent of the character of the organism's experience. What causes the confusion about what the Empiricists believed is the bad habit of reading "occasioned by experience" as entailing "learned" when, in fact, the inference is valid only in the opposite direction.

To complete our survey of the issues on which Empiricist and Nativist accounts of concept acquisition agree:

(17) Both sides assume that the space of concepts potentially available to any given organism is completely determined by the innate endowment of that organism. This follows from the assumptions that (a) the set of potentially available concepts is the closure of the primitive concepts under the combinatorial mechanisms; (b) the set of potentially available primitive concepts is innately fixed; and (c) the combinatorial mechanisms available are themselves innately specified.

I hope that, given the previous discussion, only (c) will now strike you as tendentious, and I think it's clear that (c) will have to be accepted. I won't try to set out the details, but the governing considerations should by now sound familiar: if concept learning is hypothesis confirmation, and if you need the combinatorial apparatus to specify the hypotheses that are available for the organism to confirm, then the one thing that can't be a consequence of concept learning is the availability of the combinatorial apparatus. This dilemma is conceptual, and loose talk about bootstrapping will buy you nothing. (For elaboration, see Fodor, 1975, especially chapter 2.)
concepts are primitive—and on the assumption that “sensory” is being used in even a sensory way, it must be that lexical concepts are a small subset of the ones that lexical items express. So, in particular, on the assumption that only sensory concepts are primitive—and on the assumption that “sensory” is normally constructs out of primitive concepts, and the primitive concepts are a small subset of the ones that lexical items express. So, in particular, on the assumption that only sensory concepts are primitive—and on the assumption that “sensory” is being used in even a remotely disciplined way—it must be that

Roughly, what Empiricists and Nativists disagree about is the structure of lexical concepts. For the Empiricist, lexical concepts normally have internal structure. This is because lexical concepts are normally constructs out of primitive concepts, and the primitive concepts are a small subset of the ones that lexical items express. So, in particular, on the assumption that only sensory concepts are primitive—and on the assumption that “sensory” is normally constructs out of primitive concepts, and the primitive concepts are a small subset of the ones that lexical items express. So, in particular, on the assumption that only sensory concepts are primitive—and on the assumption that “sensory” is being used in even a remotely disciplined way—it must be that

Well, then, what—insofar as Nativist and Empiricist theories of concept attainment disagree—do they disagree about? I want to answer this question twice: first roughly, by way of scouting out the logical geography; and then more precisely, by way of seeing the issue in detail.

Notice, for starters, that Nativists and Empiricists disagree on the extent to which the acquisition of lexical concepts is a rational process. In respect of this disagreement, the traditional nomenclature of “Rationalism vs Empiricism” could hardly be more misleading. It is the Empiricist view that the relation between a lexical concept and the experiences which occasion its acquisition is normally rational—in particular, that the normal relation is that such experiences bestow inductive warrant upon hypotheses which articulate the internal structure of the concepts. Whereas, it’s the

THE PRESENT STATUS OF THE INNATENESS CONTROVERSY

I am being a little cavalier but only because, in point of historical fact, this issue isn’t in dispute. It was, for example, presumably because they dreaded the sort of regress I’ve just alluded to that theorists who held that the combinatorial structure of complex concepts is associative also held that the principles of association express unlearned, innate, congenital properties of mental processing. Hume again: Association is “a kind of attraction, which in the mental world will be found to have as extraordinary effects as in the natural, and to show itself in as many and as various forms. Its effects are everywhere conspicuous; but as to its causes, they are mostly unknown and must be resolved into original qualities of human nature, which I pretend not to explain” (p. 321; Hume’s emphasis). But, now, if you hold that the primitive concepts are innate (see above), and if you hold that every concept is either primitive or complex, and if you hold that the complex concepts are the closure of the primitive concepts under some innate combinatorial apparatus, then you’d bloody well better hold that the set of potentially available concepts is innately fixed. And, of course, the Empiricists did hold that; it was precisely because their theory guaranteed the exhaustiveness of the disjunction sensory concept or complex concept reducible to sensory constituents that they were able to derive the Empirist maxim as, as it were, a theorem of their psychology. “Nothing can be in the mind except what’s first in the senses” because the potentially available conceptual repertoire is exhausted by the closure of the primitive concepts under the combinatorial apparatus.

Well, then, what—insofar as Nativist and Empiricist theories of concept attainment disagree—do they disagree about? I want to answer this question twice: first roughly, by way of scouting out the logical geography; and then more precisely, by way of seeing the issue in detail.

Notice, for starters, that Nativists and Empiricists disagree on the extent to which the acquisition of lexical concepts is a rational process. In respect of this disagreement, the traditional nomenclature of “Rationalism vs Empiricism” could hardly be more misleading. It is the Empiricist view that the relation between a lexical concept and the experiences which occasion its acquisition is normally rational—in particular, that the normal relation is that such experiences bestow inductive warrant upon hypotheses which articulate the internal structure of the concepts. Whereas, it’s the

concepts like TRIANGLE, BACHELOR, XYLOPHONE, CHICAGO, HAND, HOUSE, HORSE, ELECTRON, GRANDMOTHER, CIGAR, TOMORROW, etc. are all internally complex. The Empiricist must hold this because, by stipulation, the Empiricist view is that the attainment of non-sensory concepts involves learning the truth of a hypothesis that exhibits their internal structure. Patently, if a concept has no internal structure, its acquisition cannot be the consequence of that sort of process.

Whereas, it is open to the Nativist to hold either that (a) all or most lexical concepts have no internal structure, or (b), if they are internally structured, nevertheless the fact that they are plays no role in the explanation of their acquisition. To put the same point another way, according to the Nativist view the story that Empiricists tell about sensory concepts also holds for a wide range of non-sensory lexical concepts: viz. that they are triggered but unlearned. That Nativists do indeed hold this is clear enough from the texts. For example, Descartes argues that if the acquisition of concepts requires sensory stimulation, that is not because “these extraneous things (distal stimuli) transmitted the ideas themselves to our minds through the organs of sense, but because they transmitted something which gave the mind occasion to form these ideas, by means of an innate faculty, at this time rather than at another . . . Hence it follows that the ideas of . . . movements and figures are themselves innate in us. So much the more must be the ideas of pain, colour, sound and the like be innate, that our mind may, on the occasion of certain corporeal movements envisage these ideas” (“Notes Against A Certain Program,” quoted in Adams, 1975, p. 770).

This underlying disagreement about the structure—or otherwise—of lexical concepts has, of course, a wide range of consequences. Notice, for starters, that Nativists and Empiricists disagree on the extent to which the acquisition of lexical concepts is a rational process. In respect of this disagreement, the traditional nomenclature of “Rationalism vs Empiricism” could hardly be more misleading. It is the Empiricist view that the relation between a lexical concept and the experiences which occasion its acquisition is normally rational—in particular, that the normal relation is that such experiences bestow inductive warrant upon hypotheses which articulate the internal structure of the concepts. Whereas, it’s the
Rationalist view that the normal relation between lexical concepts and their occasioning experiences is brute-causal, i.e. "merely" empirical: such experiences function as the innately specified triggers of the concepts which they—to borrow the ethological jargon—"release."

Indeed, the ethological precedents from other species suggest that such relations may often be extremely arbitrary. There are fledgling ducks for which the operative rule seems to be: if it moves, it's mother. The fact that this experience triggers this concept is surely to be understood on a Darwinian model and not on the model of concept learning. In particular, the observation that the stimulus moves is surely not functioning as data confirmatory of some universal duckish hypothesis about the internal structure of MOTHERHOOD. The point is that the Nativist view allows for any amount of this sort of arbitrariness that you like.

Since the Nativist ipso facto doubts that ambient stimulation must provide inductive support for the concepts it occasions, it is open to him to predict the emergence of the species specific conceptual repertoire across a wide variety of nomologically possible environments. All that these environments need to have in common is that they provide the relevant triggers. It is thus a very characteristic Rationalist move—you'll find it from Descartes to Chomsky—to emphasize that the environment of the developing organism actually provides a poor inductive basis for the concepts that the organism acquires. "You learn your language on the basis of a fragmentary and impoverished sample" is one of those sentences that has been uttered more than once. The logic of this argument should now be apparent: if much of the conceptual repertoire is triggered, it's understandable that it should be invariant over wide— and relatively chaotic—variation in the individual histories of organisms belonging to the same species. To put it crudely: the concept isn't coming from the environment, it's coming from the organism. All the environment does is provide the triggers that release the information.

It is also characteristic of Rationalist theorizing to emphasize the degree to which the conceptual repertoire of the human species does seem to be largely invariant across variation in experience—for example, that the range of scientific hypotheses ever seriously entertained is vastly undetermined by the data that choose among them. From the Rationalist point of view, the problem posed by the history of science is to explain why there is so much less disagreement than there ought to be, and why such disagreements as there are are resolved by evidence for which "logically inconclusive" is an epithet that puts it mildly. I think that this is a very interesting way of looking at the history of science. I think it's probably the right way. But this is now pure digression, and I will get back to work.

I don't quite like the way that the exposition has gone so far, for it makes the dispute about concept attainment primarily a dispute about lexical concepts (the ones that are, de facto, expressed by non-phrasal predicates of English), and, as I remarked at the outset, it's reasonable to doubt that the lexical/phrasal distinction will bear much weight. Indeed, on the Empiricist account, it's an accident—a mere fact about the history of the language—that a given concept is lexical; or, if it's not an accident, it's the consequence of the functioning of such unremarkable mechanisms as Zipf's Law. Similarly, though a Nativist might reasonably take the fact that a concept is lexicalized as prima facie evidence that it's primitive—and would thus be led to predict (what seems plausible in view of the data) that the lexical/phrasal distinction should be drawn in largely similar ways across historically unrelated languages—still, it's open to a Nativist to admit that some lexical items do, in fact, express complex concepts. He could thus hold that though (pace Empiricists) the primitive/complex distinction isn't nearly coextensive with the sensory/non-sensory distinction, it may nevertheless be a fair way from being coextensive with the lexical/phrasal distinction, too. Briefly: the Nativist probably will claim that many de facto lexical concepts are internally unstructured, but that's not the heart of his position: Nativism isn't, primarily, a thesis about language; still less is it primarily a thesis about English.

So, let's kick away the ladder. We could simply put it that the dispute is over which concepts are primitive, the Nativist claim being that there are very many unstructured (hence primitive, hence unlearned) concepts which are, nevertheless non-sensory. This would be a good way of putting it except that even many Empiricists now doubt that the sensory/non-sensory distinction will survive serious scrutiny. Certainly it is unreconstructed as
things now stand, so it's not much use as a machine for clarifying other issues.

What might be closer to the point is to characterize the Nativist as denying that the primitive concepts constitute an epistemologically interesting set. From the Nativist point of view, which concepts are the triggered ones is basically a biological matter. It would be no surprise—indeed, it would be just what you'd expect—if that set looked unprincipled from the epistemologist's point of view. A Nativist would thus be quite comfortable with its including lots of middle-sized object concepts (FOOD, MOTHER), lots of explanatory concepts (CAUSES, INTENDS), as well as lots of sensory concepts, and God knows what all else.

There are, however, two things wrong with this way of putting the issue. In the first place, some Nativists, including especially many Cartesian Rationalists, have sought to make epistemological mileage out of the innate/learned distinction. Descartes agreed with Locke and Hume that the innate ideas can be picked out by an epistemological criterion; what they differed over was just which epistemological criterion was the one to use. Descartes took it that it was clarity and distinctiveness whereas, as we've seen, the Empiricist's view was that the innate ideas are the ones that are immediately occasioned by sensory stimulation. I think it was a serious mistake on the part of the Rationalists to thus muddle their epistemology with their ethology, but it seems to be the fact that muddled them they did.

The second point is that, while latter-day Empiricists continue to suspect that the primitive concepts have a special epistemological centrality (so that verificationism is epidemic in AI), many have ceased to believe that they form an epistemologically homogeneous collection. Typical contemporary formulations allow as primitive not just a sensory basis and a logical syntax, but also a heterogeneous and largely unspecified body of "framework" concepts, from which the mass of de facto lexical concepts are alleged to be constructed. (For a useful example, see Miller and Johnson-Laird, 1976, p. 14). This is a bit of a nuisance because it means that while classical Empiricism was a principled doctrine ('I'll tell you which concepts I claim are primitive: it's all the sensory ones and only those'), modern Empiricism tends to be in the nature of a research program ('The primitive concepts are the ones whose attainment I can't eventually account for by appeal to the mechanisms of concept learning. I don't yet know which those concepts are, though I very much doubt that they include XYLOPHONE or GRANDMOTHER. Ask me again later.') Viewed this way, the Nativist/Empiricist controversy is a disagreement about how this research program will turn out. The Empiricist bets that there will prove to be lots of interesting reductions of prima facie uncomplex concepts (e.g. of de facto lexical concepts); whereas the Nativist bets that the successes of reduction will be unsystematic and thin on the ground, that we aren't going to be able to display the internal structure of most concepts because, simply, most concepts don't have any internal structure.

I propose to leave it here. This formulation allows that the disagreement is, after all, largely quantitative (how many concepts will prove to be reducible?). It allows that the disagreement would be principled if the Empiricist could get his act together to the extent of deciding what precisely the epistemologically interesting characteristic of primitive concepts is supposed to be. And, most importantly, it puts the emphasis at a place where empirical considerations can be brought to bear: it puts the emphasis on the question whether, and which, concepts have internal structure. Let me reemphasize this last clause. The Empiricist story about the acquisition of concepts can, in principle, hold only for those concepts which have internal structure—those concepts which can be viewed as in some sense logical constructs out of a primitive basis in which they are not themselves contained. If it turns out that most concepts cannot be so viewed, then the Empiricist story will have to be abandoned.

The State of the Evidence

I am now going to wave my hands in lieu of providing a really systematic survey of the data. It seems to me that the burden of the empirical evidence overwhelmingly favors the view that most concepts are internally unstructured. Indeed, it seems to me to be among the most important findings of philosophical and psychological research over the last several hundred years (say, since Locke first made the reductionist program explicit) that attempts at conceptual analysis practically always fail.

Consider, for example, the failure of the reductionist program within the study of language. As we've seen, it's characteristic of the Empiricist to claim that an adequate primitive basis for the
concepts expressible in English would be much smaller than the set of de facto lexicalized concepts of English. On the further assumption that English is capable of expressing the primitive concepts themselves (an assumption you’re likely to grant if, for example, you hold that primitive concepts are sensory), you get what I’ll call the Definition Hypothesis: (a, weak version) that many de facto lexical concepts are definable; and (b, strong version) that they are definable in a vocabulary of sensory-terms-plus-logical-syntax.

It’s simply notorious that the stronger version of this claim has proved to be untenable. There are, I think, no cases where it has been possible to show that a prima facie non-sensory term (like, say, ‘table’ or ‘grandmother’ or ‘electron’ or, for that matter, ‘bachelor’ or ‘triangle’) is analyzable (i.e. definable) in a reduction base where the non-logical vocabulary is—even prima facie—sensory. That’s why the Empiricist program, which was healthy as recently as the Logical Atomist movement, is now effectively a dead issue in philosophy. But what’s equally true, and considerably more striking, is that the evidence seems to bear against the definition hypothesis even in the weak version; if there are no plausible cases of definition in a sensory vocabulary, there are also remarkably few plausible examples of definition in a non-sensory vocabulary, one indication of which is the striking paucity of working examples in the standard literature. There is ‘bachelor’, which is supposed to mean ‘unmarried man’; there are the causative verbs, of which the analysis is vastly in dispute; there are jargon terms, which are explicitly and stipulatively defined (‘ketch’: ‘a two-masted sailing vessel, such that . . .’) (See Chapman and Maloney (1977)); there are kinship terms, which are also in dispute; there is a handful of terms which belong to real, honest-to-God axiomatic systems (‘triangle’, ‘prime’ as in ‘prime number’); and then there are the other half million or so lexical items that the OED lists. About these last apparently nothing much can be done.

Not for lack of trying. The attempt to axiomatize the vocabulary of, say, ethics (i.e. by showing that you can get by with a reduction base which includes, perhaps, ‘ought’ and ‘good’ but not ‘right’) constitutes a sustained and serious contribution to the philosophical literature. The trouble is the one that John Austin pointed out (in a pregnant paper called, “Three Ways of Spilling Ink” (1966)): there seem always to be counterexamples to the proposed definitions. This observation, if true, needn’t be mysterious. Suppose that the vocabulary of English is not much larger than the minimum you would require to make the distinctions that English can encode—i.e. that, far from being vastly redundant, the vocabulary of English is just about adequate to its expressive power. Then you’d expect to find that programs of definitional elimination keep running into trouble with the examples. My claim is that that is what we do find, practically everywhere we look; and that this is inductive evidence (by affirmation of the consequent) that most of the morphemically simple expressions of English are undefinable. (I once heard Professor Gilbert Harman remark that it would be surprising if “know” were definable, since nothing else is. Precisely.)

The only way to convince yourself of the truth—or otherwise—of this claim is to try actually working through some cases. While we can’t do that in any systematic way here, it may be worth looking at one example (chosen more or less arbitrarily from the current literature) just to give a little of the feel of the thing.

In a recent (1978) paper, in which the definition construct is heavily exploited, George Miller makes the following claim (p. 104): “When nouns of Type M are used as verbs, the meaning of x Ms y is to be construed as ‘x covers the surface of y with M’.” Miller takes the noun ‘paint’ as his paradigm of the type M, though he mentions a number of others as well: butter, color, dye, enamel, grease, etc.

What we have here is a proposal for defining the transitive verb ‘paint’ in terms of the noun ‘paint’, together with some further conceptual apparatus (COVER, Surface, and WITH), and what I claim is that the definition doesn’t work; x covers y with paint may be a necessary condition for x paints y but it is certainly not a sufficient condition. (It is, by the way, quite usual for putative definitions to fail in this particular way; it is very much easier to provide logically necessary conditions for the application of a term than to provide logically necessary and sufficient conditions.)

The first point is that the reason that the definition doesn’t work has nothing to do with whether the verb ‘paint’ is vague, though that’s a red herring that often swims by about here. For example, someone might say ‘x covers the wall with paint’ isn’t
necessary for 'x paints the wall' since, surely, you could have painted the wall even if there were a little tiny bit up there in the top left hand corner that you'd missed, hence failed to cover with paint. In that case, it might be that you'd painted the wall carelessly, or imperfectly, or not very well; but, nevertheless, you'd have painted it.

But the (presumptive) fact that 'paints' is vague is not a reason for believing that it is indefinable; all it shows is that whatever phrase defines 'paints' must also be vague, and in the right ways. In fact, the definition in terms of 'cover' does the job quite nicely; for if you can be said to have painted a wall even though you missed a spot, you can also be said to have covered the wall with paint even though you missed a spot. The vagueness of definiens and definiendum, as it were, match; and so long as they do so, vagueness is not an argument against definability.

What's wrong with the definition isn't that 'paints' is vague; it's that there are clear counterexamples to the coextensivity of 'x paints y' with 'x covers y with paint.' To start with a fairly crude point, consider the case where the paint factory explodes and covers the spectators with paint. This may be good fun, but it is not a case of the paint factory (or the explosion) painting the spectators. What we need to patch up the definition is at least the condition that what goes in for 'x' must denote an agent. I don't know what an agent is (nor does anybody else), but it must at least be such as to include John in 'John painted the wall' and to exclude the paint factory in 'the paint factory exploded and covered the spectators with paint.' Notice, by the way, that in adding this condition, we are enriching the presumed reduction base for 'paints'—hence increasing the range of concepts that our theory about 'paints' claims that a child has to have mastered before he can master that verb. It goes without saying that the more such apparatus we have to add to the reduction base for defining 'paints', the more elaborately nativist is the theory which says that learning the verb is learning the definition. In the limiting case, we say that the concept PAINT_TR belongs to the reduction base for the verb 'paints', at which point the theory that learning the verb is learning its definition is equivalent to the theory that the concept PAINT_TR is innate, and argument ceases.

It might be argued, however, that conceding AGENT in the reduction base for 'paints' isn't really so bad since, when you come to think of it, AGENT is a very abstract concept and a very general one, and is presupposed by the idea of an action, which itself turns up as a constituent of lots and lots of the concepts that children have to master. So perhaps we should live with presupposed unanalyzed AGENT and revise the definition of 'paints' to read something like 'x paints y iff x is an agent and x covers the surface of y with paint.' The trouble is that this definition doesn't work either.

Consider that Michelangelo, though an agent, was not a housepainter. In particular, when he covered the ceiling of the Sistine Chapel with paint, he was not painting the ceiling; he was painting a picture on the ceiling. Similarly, when Jackson Pollock flung the pigments about, he wasn't painting his canvases; he was painting a picture ('making an image'?) on the canvases. Compare Tom Sawyer and his fence.

What's gone wrong here is maybe this: it's not good enough that you cover the y with paint. It's not even good enough that you cover the y with paint intentionally (though that's clearly necessary since, if you knock over the paint bucket, thereby covering the surface of the floor with paint, you have not thereby painted the floor). It's got to be that when you cover the y with paint, what you primarily have in mind to do (the description, as it were, under which you intend your act) is that y should be covered with paint in consequence of your activity. If, like Michelangelo, what you primarily have in mind is not (just) that the surface should become covered with paint in consequence of your activity, but that there should be a picture on the surface in consequence of your activity, then what you're doing when you cover the surface with paint doesn't count as painting the surface.

So perhaps we should try something like: 'x paints y' means 'x is an agent and x covers the surface of y with paint, and x's primary intention in covering the surface of y with paint was that the surface of y should be covered with paint in consequence of x's having so acted upon it.' I pause to remark (a) that the concept PRIMARY INTENTION OF AN ACT is quite unanalyzed; (b) that it is vastly unlikely that children have access to that concept prior to their learning of the concept PAINT_TR; and (c) that any way you look at it, the definition of 'paints'—the amount of subsidiary
apparatus you need to define it in terms of ‘paint’—is getting sort of hairy.

Anyhow, this definition doesn’t work either. For consider that when Michelangelo dipped his brush into Cerulian Blue, he thereby covered the surface of his brush with paint and did so with the primary intention that his brush should be covered with paint in consequence of his having so dipped it. But MICHELANGELO WAS NOT, FOR ALL THAT, PAINTING HIS PAINTBRUSH. (He was just putting paint on his paintbrush.)

I don’t know where we go from here. For all I know—for all anybody knows—‘paint’ is undefinable; for all I know, you can’t eliminate it even in terms of such a very closely related term as ‘paint’. Or perhaps it is definable, but only in a reduction base that includes ‘dinosaur’ and ‘Chlorodont’. Either way, the present point is that Miller’s example doesn’t work. That’s not surprising; when it comes to definitions, the examples almost always don’t work.

Now consider the case of the definition construct psycholinguistics. If you think that most lexical concepts are complex, you presumably suppose that entertaining a complex concept is normally part and parcel of comprehending (token) utterances in your language. In particular, you will assume this even in the case of utterances of morphemically simple forms, so long as the monomorpheme uttered is definable. A fair amount of experimental work has recently been devoted to testing this assumption—for example, by trying to show that the relative complexity of the state of mind you are in when you grasp a prima facie definable expression is predictable from the relative complexity of the definition. I think it’s fair to say that this enterprise has been quite uniformly unsuccessful. (For discussion, see Fodor et al., 1980.)

There is some evidence that the state of mind you’re in when you hear an expression is sometimes in a certain sense close to the one you’re in when you hear a term that figures in its definition; e.g. that when you hear “cat” you’re in a state of mind closely related to the one you’re in when you hear “animal” (see particularly Meyer and Schvanteveld, 1971). What’s lacking are any empirical grounds for claiming that the latter state of mind is, as it were, a constituent of the former: that understanding “cat” involves entertaining ANIMAL. We’ll see presently that this point is not novel; the lack of evidence for such constituency was noticed relatively early by Empiricist theorists, and it occasioned something of a crisis in the history of Associationism.

To put it briefly: when viewed from the perspective of the theory of language, the idea that many terms express concepts that have internal structure is tantamount to the idea that many terms have definitions. This latter doctrine is abstract but empirical: the relevant demonstration would involve exhibiting the alleged definitions and showing that they do play a role in the psychological processes that mediate the use of language. I’m claiming that—ever so many texts to the contrary notwithstanding—the evidence for this doctrine is not good; that, on balance, the attempts to apply the program of conceptual reduction to the study of the lexicons of natural languages have not been fruitful.

It would, however, be a serious mistake to attempt to evaluate the empirical status of that program solely by reference to its application in lexicography and psycholinguistics. On the contrary, what seems to me impressive—what ought to worry an Empiricist—is the range of enterprises in which conceptual reduction has been tested and, so far at least, found wanting.

I shall scarcely do more than mention the failure of such undertakings in Positivistic philosophy of science. Having convinced themselves that the warrant of inductively founded beliefs turns on the possibility of reducing the theoretical concepts of a science to its observation base, the Positivists undertook a sustained and extremely ingenious attempt to provide—or at least to schematize—the required reductions. It is now, I think, practically universally conceded that this attempt was unsuccessful and, indeed, cannot in principle be carried through. Philosophers disagree about why, precisely, the Positivist program failed; but I doubt that anyone now believes that concepts like ELECTRON are fruitfully viewed as constructs out of an epistemologically specifiable primitive basis.

What is perhaps closer to the heart of a cognitive scientist is the vicissitudes of the program of conceptual reduction in its application to theories of perception. It’s usual, in the context of representational theories of mind, to think of perception as the application of a concept to one of its instances. If you think that most concepts are constructs out of a (more or less) sensory basis,
then a fortiori you think that sufficient conditions for the application of a concept can be specified in (more or less) sensory terms. Perception can then be viewed as a process in which the organism determines whether the sensory criteria for the application of a concept are satisfied in a given stimulus array. On this view, the internal structure of a concept determines the flow of information through the perceptual mechanisms when the concept is applied: If the concept is conjunctive, perceptual processing is the search for stimulus properties which satisfy each of the conjuncts; if it is disjunctive, perceptual processing is the search for stimulus properties which satisfy either of the disjuncts, and so on.

This determination of information flow by conceptual structure is, of course, only very approximate, the possibility of much heuristic shortcutting being conceded on all hands. Nevertheless, in point both of logic and historical fact, people who hold the reductionist view of conceptual structure typically do and ought to hold a basically bottom-to-top view of perceptual analysis. In the limiting case, the internal structure of a concept tells you what you have to know—and all you have to know—to determine whether the concept applies. So in the limiting case the application of a complex concept ought to reduce to the determination of the applicability of its primitive constituents. The flow of information in perception ought, fundamentally, to go in one direction: from the application of sensory concepts, via processes of logical construction, to the application of complex concepts. (Compare the Empiricist theory of confirmation: certitude flows upward, via logical construction, from the observation base of a science to its theoretical sentences; the logical constructions involved are those exhibited by reductive definitions of the theoretical concepts).

My point is that, insofar as the Empiricist view of concepts is wedded to the bottom-to-top view of perceptual processing, the quite general failure of research programs based on the latter doctrine is ipso facto an argument against the former. The relevant consideration is not just that it's hard to specify an adequate sensory basis for perceptual integration, to specify logically sufficient sensory conditions for the application of perceptual concepts. What's more important is that information seems to flow in every direction in normal perceptual processing. For all that the study of perception shows us, every concept could be a primitive node in a conceptual network, and any connection between nodes might be exploited, at one time or another, in constraining hypotheses about which concepts are satisfied in the stimulus domain. The analogy between theories of perception and theories of confirmation is once again instructive: empirical warrant may be inherited in any direction through a network of empirical theory. It's as likely that we accept a datum because it comports with abstract principle as that we accept an abstract principle because it comports with our data. That, as Hume pointed out, is why we don't believe reports of miracles.

I think one could go on to make very similar points about the way that the internal structure story has come to grief in the study of conceptual development in children. But I won't; I now propose to stop waving my hands. Quite a useful book might be written about how the idea that concepts have internal structure has determined the shape of several centuries of research in philosophy, psychology, lexicography, and, most recently, in AI. My own view is that the moral is surprisingly clear: all such programs have failed, and they have failed in just the respects you would expect them to if the doctrine that concepts have internal structure is untrue. But I don't suppose that the preceding fragments will have convinced anyone of this, and I'm content to settle for mere willing suspension of disbelief. In what follows, I'll assume that the evidence is, in any event, not overwhelmingly in favor of the conceptual reduction program and that it might, therefore, be interesting to consider some alternatives. In particular, I want to look at two doctrines—one modern and one traditional—which are prepared to abandon some aspects of the standard reduction account, but which seek to block the inference from the failure of conceptual reduction to the Rationalist view of concept attainment. I'll argue that while both these doctrines have interesting lessons to teach, neither succeeds in its primary objective: upon examination, one of them proves to be a kind of closet Rationalism, and the other offers a theory of concepts which is inadequate on grounds independent of the ones that Empiricists and Rationalists disagree about. I shall consider these proposals in reverse chronological order. At the end I'll say a little—a very
be learned: undefinable concepts with prototype structure could be learned by (roughly) any device that can estimate central tendencies.

Now, what is striking about prototypes as opposed to definitions is that, whereas the evidence for the psychological reality of the latter is, as we've seen, exiguous, there is abundant evidence for the psychological reality of the former. Eleanor Rosch (1975) and her colleagues, in particular, have provided striking demonstrations that the prototype structure of a concept determines much of the variance in a wide variety of experimental tasks, chronometric and otherwise. So, for example, you can estimate the prototype structure of a concept by asking subjects for the first instance of that concept which occurs to them. (If I say: “Tell me the first kind of gem you can think of” you will probably say “diamond” and you probably will not say “topaz”. Similarly, mutatis mutandis, if I say: “Tell me the first property of diamonds you can think of” you will probably say “hard” or “expensive” and you will probably not say “treasured by Armenians”; not even if you believe that Armenians do, in fact, treasure diamonds.) It turns out that such estimates of prototype structure reliably predict speed and/or accuracy of performance in a variety of experimental tasks like:

—say “true” or “false” to “a diamond is a gem”. (“Diamond” is faster than “topaz”).
—say “true” or “false” to “diamonds are hard”. (“Hard” is faster.)

Similarly, estimates of its prototype structure reliably predict the relative frequency of lexical items which express a concept (“diamond” is more frequent, in English, than “topaz” is). They reliably predict order of lexical acquisition (children learn “diamond” before they learn “topaz” (Anglin, 1970); and they surely learn that diamonds are hard before they learn that Armenians like them.) Estimates of prototype structure reliably predict some of the facts about ‘semantic’ priming (a superordinate concept differentially primes its high-prototypicality instances). And so on.

Insofar as theses get established in cognitive psychology, I think we can take the reality of prototype structures as read. The
question is: what does this fact show us about the question of concept attainment? One view, currently much in the air, (see Putnam (1975)) goes like this:

Many de facto lexical concepts are really quasi-indexical. In particular, what you actually learn when you learn a concept like TIGER is that a tiger is an animal of the same kind as this, where the "this" indicates an ostensive introduction. That sort of mental representation is able to function in (e.g.) mediating communication with the word "tiger" and in the perceptual recognition of tigers because a psychological counterpart to the notion of same kind is provided by a tiger prototype. You learn this prototype when you acquire the concept and you share it, more or less, with other members of your speech community—that is, what the ostensively introduced tiger has in common with other tigers, insofar as what tigers have in common is internally represented by people who have the TIGER concept, is just that it's a reasonable fit to the prototype. What is not internally represented—or, anyhow, what needn't be—is the defining properties of tigers; properties which all and only tigers share in virtue of being tigers. Those properties—if there are any—have to do with biological or genetic structure and are matters for scientific discovery in zoology, not in psychology.

I'm inclined to think there is more than a grain of truth in this (Putnam/Kripke) account, in particular that the proposal that natural-kind concepts are often quasi-indexical is correct and important. What I want to consider here, however, is whether the idea that concepts have prototype structure—that what are in the head are prototypes rather than meanings—will do the work that a theory of concepts needs to do.

Notice, to begin with, that in principle the claim that kind concepts are quasi-indexical comes unstuck from the claim that what's in the head doesn't determine extensions. After all, if what I'm pointing at is a tiger, then the formula "... is of relevantly similar kind to what I'm pointing at" does, in fact, determine the extension of "tiger": all and only the tigers satisfy that formula. What makes the Putnam/Kripke story incompatible with the idea that concepts determine extensions isn't the stuff about their being quasi-indexical; it's the stuff about their having prototype structure.

Second, the experimental evidence that prototype structure is psychologically real certainly isn't incompatible with the claim that mental representations determine extensions; it isn't even incompatible with the claim that mental representations have the form of definitions. As several investigators (Armstrong et al., forthcoming; Wanner (1979)) have recently remarked, concepts like GRANDMOTHER and TRIANGLE, which are definable concepts par excellence, and where, often enough, the subject can tell you what the definition is, nevertheless exhibit the characteristic range of prototype phenomena. (The prototypical grandmother is elderly and has white hair; the prototypical triangle is equilateral, etc.) Since prototype structure can, patently, comport with definitional structure, the inference from many concepts have prototype structure to many concepts are complex though undefinable is, at best, risky.

'Still,' you may say, 'the prototype idea shows how concepts could be complex but undefinable, and this lets the wind out of the Rationalist's sails. In particular, we now see that evidence for its undefinability just won't do to make the case that a concept is primitive; and it's primitiveness you need if you are to infer that a concept is unlearned. 'And anyhow,' you may add, in one of your Wittgensteinian moods, 'doesn't it seem plausible that the internal representation of a word is typically more or less of a mess? Isn't what games and tigers and typewriters have in common, each with each, more like a family resemblance than like the mutual satisfaction of defining properties? Isn't a language really sort of like a city? Or a rope? Or a fountain? Or something?'

Well, maybe. But there is a principled argument against the prototype story, and unless or until that argument can be met, I see no way of taking prototypes as serious candidates for identification with concepts-qua-mental-representations. The objection goes like this:

The classical mentalist view is that postulating concepts answers three questions: (a) what does a term in a natural language express; (b) what determines the extension of a term in a natural language; and (c) what does a term in a natural language contribute to the semantics of the complex expressions in which it occurs. The beauty of the classical story is that it provides the same answer for all three of these questions: in particular, the very thing which determines the extension of a term is also what the term
contributes to the (syntactically) complex expressions that contain it. So, for example: the concept COW determines the extension of “cow”. “Cow” expresses COW, and it contributes the concept it expresses to, for example, the internal representation of “brown cow”. Since “brown cow” inherits COW from “cow” (and, of course, inherits BROWN from “brown”), and since concepts determine extensions, it’s possible to compute the semantic properties of “brown cow” from the semantic properties of its constituent terms. It is because it’s possible to compute the semantic properties of complex expressions from the semantic properties of their constituent terms that it is possible to understand an infinite language given knowledge of a finite basis. It’s worth adding that this general account is not proprietary to the mentalistic reading of “concept”; the Fregean program in semantics runs along precisely the same line.

So, in short, the idea that concepts determine extensions is part and parcel of the classical theory of how the productivity problem is to be solved. Moreover, I find it hard to imagine a solution to the productivity problem that doesn’t share the form of the classical one; one that doesn’t say that there is productivity because semantic properties are inherited from terms to phrases (or, equivalently for our purposes, from lexical concepts to phrasal concepts). Contrapositively, nothing could be the meaning of a word unless it’s what phrases in which the word occurs inherit from the word’s occurrence. The anti-prototype point is just that if that generalization stands, then prototypes can’t be word meanings.

The argument is blessedly short: the only obvious way in which prototypes could be inherited under combination would be for the prototypes of complex expressions to be determined by the prototypes of their constituents (just as the meanings—hence extensions—of complex expressions were classically supposed to be inherited from the meanings of their constituents). But you can’t play that game with prototypes because, by and large—i.e. in infinitely many cases—complex expressions don’t have prototypes. There may, for example, be prototypical cities (London, Athens, Rome, New York); there may even be prototypical American cities (New York, Chicago, Los Angeles), but there are surely no prototypical American cities situated on the East Coast just a little south of Tennessee. Similarly, there may be prototypical grandmothers (Mary Worth) and there may be prototypical properties of grandmothers (good, old Mary Worth). But there are surely no prototypical properties of, say, Chaucer’s grandmothers, and there are no prototypical properties of grandmothers most of whose grandchildren are married to dentists. Or, if there are, it’s clear that knowing the prototypes can’t be required to understand the phrases.

Since phrases typically do not have prototypes, it cannot be that the combinatorial inheritance of semantic properties is a matter of the determination of phrasal prototypes by lexical prototypes. And it cannot be that understanding a phrase is a matter of knowing how its prototype is determined by those of its constituents. And since a serious theory of concepts must say what lexical items contribute to phrases, I infer that the prototype story is not a serious theory of concepts.

Or not, at least, so far. For it is, of course, possible that the combinatorial problem could be solved in some way compatible with the idea that lexical concepts are prototypes—i.e. in some way other than the one I have just sketched. It’s unfortunately highly characteristic of the neo-Wittgensteinian line in lexicography that the problem of squaring the prototype story with the demands that the problem of linguistic productivity imposes simply isn’t broached. Indeed, it’s highly characteristic of these texts, from the Investigations to “The meaning of meaning” inclusive, not to mention the productivity problem at all. My point has been that lexicography so pursued is lexicography unconstrained.

It is worth pausing to block a possible misunderstanding of the argument just proposed. There exist formal mechanisms which might, at least in principle, permit prototype theories to handle the combinatorial problem for the extensions of terms. Indeed, it is a standard assumption among many friends of prototypes that some variant of “fuzzy set theory” could do the job: “cow” expresses a fuzzy set, “brown” expresses a fuzzy set, and the combinatorial mechanisms exhibit “brown cow” as expressing their fuzzy intersection. There are serious problems about characterizing these combinatorial mechanisms (see Osherson and Smith, in press) but that is not the difficulty that I want to emphasize here. The present problem is to give some account of the sense (as op-
posed to the extension) of "brown cow," assuming that the senses of "brown" and "cow" are taken to be prototypes. It appears that this problem could be solved only if prototypes have some sort of combinatorial structure; and, so far as anybody knows, they don't.

To return to the main discussion, my point hasn't been, of course, that there are definitions rather than prototypes. The current state of the argument is this:

- Rationalists say that lexical concepts are typically undefinable, hence typically unstructured, hence typically primitive, hence typically unlearned.
- The Empiricist counters: you can't infer from undefinable to unstructured. Perhaps lexical concepts typically have prototype structure.
- The Rationalist replies: if prototypes offered a serious account of conceptual structure, then the inference from undefinable to unstructured would indeed be jeopardized. You have thus shown, at least, that the inference was apodictic (or, if I did, I shouldn't have). My argument was (or should have been) just that there's no evidence for conceptual structure and there's some evidence (apparent undefinability) against it. If there is no structure inside lexical concepts, that would explain their apparent undefinability. So far, no other plausible explanation suggests itself. In particular, the explanation that goes "lexical concepts are complex but do not determine extensions" apparently won't do.

And there, for the moment at least, the argument comes to rest.

In intellectual history, everything happens twice, first as philosophy and then as cognitive science. It turns out that several of the points in the preceding discussion were anticipated by the "mental chemistry" movement in British Associationism. I want to comment on this movement, since the ideas are of considerable intrinsic interest and some of their implications appear to have been widely misunderstood, even by the philosophers who espoused them.

I think many readers may be laboring under a minor terminological misapprehension of which I was myself only recently dis-
extend—in particular, to the fact that introspection—then the preferred technique for the empirical investigation of mental states—often fails to find psychological structure in concepts in which semantic analysis discovers logical complexity. The germ of the problem, and of the proposed solution, can be found quite early in the Empiricist literature. So, for example, David Hartley—who is approximately a contemporary of Hume’s, writes as follows:

If the number of simple ideas which compose the complex one be very great, it may happen, that the complex idea shall not appear to bear any relation to those its compounding parts, nor to the external senses upon which the original sensations, which gave birth to the compounding ideas, were impressed. The reason of this is, that each simple idea is overpowered by the sum of all the rest, as soon as they are all intimately united together. Thus, in very compound medicines, the several tastes and flavors of the separate ingredients are lost and overpowered by the complex one of the whole mass: so that this has a taste and flavor of its own, which appears to be simple and original, and like that of a natural body.” (Hartley, 1966, p. 75)

Notice two things: (a) the distinction between the properties that philosophical analysis predicts that a concept ought to have and the ones that psychological (viz. introspective) analysis finds that it does have; and (b) the proposal that the disparity be explained by appeal to some form of interaction among simple ideas other than mere summation (other, that is, than what we’ve been calling logical construction). These are the leitmotifs of mental chemistry.


The way out of this difficulty appears plain enough to the younger Mill. It involves merely abandoning the mechanical view of association for the chemical analogy. According to him, association produces experiences which are really unitary, in the sense that they are not separable into parts by introspection. Introspection does not break the experience up into elements... but it does reveal the elements which the act of association has brought together to form this new product. Association does more than unite—it transforms... what James Mill considered ingredients in a complex experience are really its antecedents... the resulting experience is not so much a complex state as a derivative state, which is introspectively as simple as it appears. Thus the perception of an orange or the concept HORSE are really simple, though derived from primitive elements.

This may all seem a little quaint, since latter-day Empiricists will surely say that what Mill ought to have done is hold onto both the logical and the psychological complexity of HORSE and give up the infallibility of introspection. As Warren points out, Mill seems to have viewed his problem as precisely a paradox of infallibility: if introspection always delivers truths, how can the results of analysis show that introspectively simple concepts are in fact complex? Or, to put it another way, one has no right to speak, as Hartley does, of a complex concept as merely appearing to be simple; the appearance/reality distinction doesn’t apply to the objects of an infallible faculty. As Warren remarks, “if introspection declares a certain experience to be simple, how can another introspection prove this introspection to be false?”

But, in fact, the introspection issue is—as usual—a red herring, and Mill’s problem is very much our own. If introspection isn’t always right, it isn’t always wrong either; and, in the present case, the current psychological evidence tends to support the earlier intuitions. There is, if anything, rather less reason to accept the psychological complexity of HORSE now than there was when Mill wrote his Logic.

What’s interesting is that Mill, having once taken the introspective data seriously, was able to see that they called for a major revision of Empiricist theorizing. And what’s still more interesting is that even Mill apparently failed to see how much taking these data seriously was going to cost him. Here is the relevant passage from Mill’s System of Logic, (1974, Vol. 3, pp. 853-854):

the effect of concurring causes is not always precisely the sum of the effects of those causes when separate, nor even always an effect of the same kind with them... the laws of the phenomena of mind are sometimes analogous to mechanical, but sometimes also to chemical laws. When many impressions or ideas are operating in the mind together, there sometimes takes
place a process of a similar kind to chemical combination... the ideas sometimes melt and coalesce with one another and appear not several ideas but one; in the same manner as, when the seven prismatic colours are presented to the eye in rapid succession, the sensation produced is that of white. But in this last case it is correct to say that the seven colours when they rapidly follow one another generate white, but not that they actually are white; so it appears to me that the Complex Idea, formed by the blending together of several simple ones, should, when it is really simple (that is, when the separate elements are not consciously distinguishable in it) be said to result from, or be generated by, the simple ideas, not to consist of them. Our idea of an orange really consists of the simple ideas of a certain colour, a certain form, a certain taste and smell, etc., because we can, by interrogating our consciousness, perceive all these elements in the idea. But we cannot perceive, in so apparently simple a feeling as our perception of the shape of an object by the eye, all those multitudes of ideas derived from other senses, without which it is well ascertained that no such visual perception would ever have had existence; nor, in our idea of Extension, can we discover those elementary ideas of resistance, derived from our muscular force, in which it has been conclusively shown that the idea originates. Therefore these are cases of mental chemistry: in which it is proper to say that the simple ideas generate, rather than that they compose, the complex ones.

There are, I think, two things that need to be said about this. We need to note the magnitude of the concessions, which amounts— as I shall argue—to abandoning Empiricism entirely; to giving the Rationalist everything he could possibly want. And we should also note that the kind of Rationalism that Mill is (implicitly, to be sure) espousing is actually quite interesting and points in a direction that modern psychologists might well want to explore. These points in turn.

To begin with, it’s clear that Mill’s proposal tacitly abandons the demand that the logical and psychological tests for conceptual structure should be satisfied by the same cases. On Mill’s account, psychologically simple concepts might nevertheless have semantic entailments; indeed, they might even have definitions (in the sense of logically necessary and sufficient conditions). Definability does not guarantee complexity because, in point of their psychological functioning, the concepts which figure in a definition may be causal occasions—not structural constituents—of the concept they define. So, in espousing mental chemistry, Mill has given up the principle that psychological constituency explains semantic entailment. All that seems transparent enough; what may not be immediately obvious is that he has given up the Empiricist Maxim as well.

Strictly speaking, what the Empiricist Maxim says isn’t, of course, that there’s nothing in the mind except what’s first present to the senses; Locke knew perfectly well about unicorns. Rather, the Maxim says that there is nothing in the mind except what’s either first present to the senses or logically complex and constructed from sensations. Equivalently, for our purposes, it says that all simple ideas must be learned from their instances. Empiricist analysis in its iconoclastic mood typically used the Maxim in the disjunctive form: take a concept (GOD, SUBSTANCE, HIDDEN POWER, or whatever), show that it is neither instantiated in experience nor reducible to concepts so instantiated, and conclude that the concept is empty.

Now, one way to put my point is that if mental chemistry is endorsed, this strategy isn’t going to work. On the mental chemistry view, the disjunction isn’t exhaustive, since there can perfectly well be concepts which are both simple and not learned from their instances. Indeed, on Mill’s account, that could be true of any simple concept; the fact that a simple concept is instantiated in experience is no reason at all to suppose that its instantiations play a role in the learning. Consider the concept C which, by assumption, is simple and arises by mental chemistry from (hence is not a construct out of) concepts A and B. (A and B may or may not be semantically related to C. The acquisition of a concept by mental chemistry—unlike its acquisition by logical construction—is neutral on the issue of logical connections between the occasioning concepts and the concept acquired.) The point is that if entertaining A and B is causally sufficient for acquiring C, then it follows that C’s being simple does not require that C be learned from its instances. All that’s required to learn C is the previous
acquisition of A and B. This is to say, in effect, that some simple concepts may be triggered by other concepts, which latter may—but need not—they themselves be simple. Simple concepts which arise as the effects of such triggering are, no doubt, learned in consequence of experience; experiences are—directly or indirectly—among their causes. But they aren’t learned from experience, and they therefore constitute exceptions to the Empiricist Maxim.

It is, by the way, of some interest that the possibility of this sort of exception to the principles of Empiricism was noticed by, of all people, David Hume, who suggested the following example:

Suppose . . . a person to have enjoyed his sight for thirty years, and to have become perfectly acquainted with colors of all kinds except one particular shade of blue, for instance, which it has never been his fortune to meet with. Let all the different shades of that color, except that single one, be placed before him, descending gradually from the deepest to the lightest; it is plain that he will perceive a blank, where that shade is wanting . . . now I ask, whether it is possible for him, from his own imagination to supply this deficiency, and raise up to himself the idea of that particular shade, though it had never been conveyed to him by his senses? I believe there are few but will be of the opinion that he can; and this may serve as a proof that the simple ideas are not always, in every instance, derived from the correspondent impressions; though this instance is so singular, that it is scarcely worth observing, and does not merit that for it alone we should alter our general maxim. (1963, p. 21)

Hume pretty obviously wasn’t much worried about this sort of case, and I think it’s clear why he wasn’t. Hume took it for granted that even if not quite all sensory concepts are learned from experience, still all non-sensory are eo ipso complex. It’s only when you put the idea that some simple concepts may not be learned from experience together with the idea that very non-sensory concepts (like HORSE or ELECTRON) might be simple that you get a mechanism capable, in principle, of generating devastating counterexamples to the Empiricist program.

Indeed, if you put the mental chemistry idea together with the idea that any (de facto lexical) concept might be simple, you get the total abandonment of the Empiricist program. What you’re left with is an account of concept attainment that is compatible with even the more ardent Nativism—for example, with Descartes’ view of concept attainment.

The Cartesian story, you’ll remember, is that even logically complex (e.g. definable) concepts like TRIANGLE are typically acquired in consequence of processes of stimulus triggering. But, as we’ve just seen, this account can perfectly well be told couched in Mill’s language of mental chemistry. I proceed to tell it so couched: "There are experiences which causally occasion the attainment of TRIANGLE. These often include (though perhaps they needn’t) seeing things that are roughly triangular; they also often include (though perhaps they needn’t) the prior attainment of such relatively elementary concepts as LINE, ANGLE, FIGURE, and the rest. Even, however, if the acquisition of TRIANGLE is causally contingent upon the prior acquisition of these elementary concepts, still the concept TRIANGLE may be psychologically simple. This is because, by assumption, it arises from the prior concepts by mental chemistry, not by construction. The fact that getting TRIANGLE requires previously getting LINE—if it is a fact; which it probably isn’t—is thus a brute fact, and it’s an accident that it corresponds to the definability of TRIANGLE in terms of LINE. For all that the Cartesian—or the mental chemistry—account requires, the brute fact might just as well have been that getting TRIANGLE is causally dependent upon first getting ELEPHANT. If getting TRIANGLE depends on first getting LINE, that is not because observations of lines occur in the induction base for hypotheses about the definition of TRIANGLES; it’s because the principles of mental chemistry are such that experiences of lines trigger the availability of LINE and the availability of LINE triggers the availability of TRIANGLE."

To say that it’s a brute fact that TRIANGLE is ontogenetically dependent upon LINE (or, for that matter, upon ELEPHANT) is to emphasize that mental chemistry—unlike “concept learning”—requires no rational relation between a concept and its causes. It isn’t, of course, to say that it’s an uninteresting fact, and it certainly isn’t to say that it’s an inexplicable fact. What, on the mental chemistry view, explains the contingency of a concept upon its causes is, of course, the principles of mental chemistry, which is to say, the innate structure of the mind.

If this last equivalence (of the principles of mental chemistry
with the innate structure of the mind) isn’t obvious, that may be because we have thus far followed Mill in avoiding saying anything about what the principles of mental chemistry are. It’s a bit curious—and it’s a fact that Mill observes in the Logic—that while mental chemistry theorists paid quite a lot of attention to the nature of association, there’s just nothing in the literature about what kinds of “fusions” of concepts mental chemistry allows. However, even the briefest consideration of what such a theory would be like suffices to show how profound the nativistic commitments of the idea actually are.

To begin with, the analogy to molecular bonding is really quite misleading; a half-hearted attempt to have it both ways. The point about molecules is that you can’t (or couldn’t in Mill’s day) predict the properties of compounds from the properties of their elements; whereas, Mill’s point about putatively complex ideas is explicitly that they don’t have elements, that they are psychologically simple. So, then, what a theory of mental chemistry must do is determine the properties of concepts as a function of the properties of (not their elements but) their causes.

We can think of the principles of such a theory as determining a mapping from, say, concepts available onto concepts attained; in particular, we can think of the theory as enumerating the prior mental states that are causally necessary and sufficient for triggering each concept that mental chemistry gives rise to. If the theory is true, then the mind is a realization of that function. The fact that the mind is structured in the way it is explains the (brute) fact that the availability of some concepts automatically gives rise to the availability of others, just as it’s the way the mind is structured which explains the brute fact that sensory stimuli trigger the sensations that they do, or that associations form on the occasions when they do. What Mill ought to say about the principles of mental chemistry is precisely what we’ve seen Hume did say about the associative principles: they are “original qualities of human nature.”

Mill’s Rationalism is historically interesting because it illuminates how short the path is from the idea that most de facto lexical concepts are simple to the idea that most de facto lexical concepts are innate. But it is also interesting because it’s an interesting kind of Rationalism. I spoke of a theory of mental chemistry as a function from (triggering) concepts available onto (triggered) concepts attained. I now want to emphasize that there is nothing in the idea of mental chemistry—or, for that matter, in Mill’s exposition of the idea—which constrains the kinds of concepts that can enter into the sorts of triggering relations that a mental chemistry specifies. In principle, mental chemistry could determine such relations among quite abstract concepts—for example, ones whose relation to experience is mediated by long chains of theory. I think this is an important idea, since any form of Rationalism must take into account the appearance of cumulation in the development of a conceptual repertoire, both in ontogeny and in the history of science; and the notion that some concepts trigger others may provide a means of doing so.

If you take triggering to be a relation between stimuli and concepts, and if you assume that practically all de facto lexical concepts are uncomplex, then you seem to be committed to saying that it’s an accident (either of the history of the individual or of intellectual history) that we get our concepts in the order that we do. For example, you seem to have to say that, given the right distribution of impinging stimuli, children might perfectly well acquire ELECTRON before they acquire NOSE (or, indeed, before they acquire any middle-sized object concepts). Similarly, you seem to have to say that it’s only an artifact of the arrangement of the triggers that we developed our geometry before our topology, or that mechanics was around before the quantum theory.

Now, of course, it is true that priorities in the development of the conceptual repertoire to some extent reflect arbitrary, local arrangements of the ambient stimulation. The concept SNOW is, presumably, available to children in Lapland earlier than it’s available to those in Africa. And it’s possible to believe that the history of astronomy would have been quite different if, on the one hand, the Earth had been the only planet in the Solar System (no data about retrograde motions) or if, on the other hand, the stars had been close enough to exhibit visible parallax. The problem is that it’s hard to believe that all the facts about the order of development of conceptual repertoires are like these. It seems plausible that some such facts would hold over quite a wide range of possible environments because they in some sense reflect dependencies among the concepts acquired. The empiricist has a model for the existence of such dependencies, since he allows
that some concepts are constructed out of others. Whether his model predicts the right dependencies is, no doubt, an open question: but at least it has parameters that accommodate the phenomenon in principle. If concept C is a construct out of concept A, then presumably the acquisition of C cannot precede the acquisition of A either in ontogeny or in the history of science.

But what if you’re a Nativist and thus believe that definition and logical construction play a relatively small role in the specification of concepts? Can we imagine mechanisms for accounting for a “normal” order of attainment even among concepts that are viewed as primitive and hence unstructured? Or, to put the question slightly differently, can we imagine principles that determine priorities in the acquisition of a conceptual repertoire other than by appeal to the facts of semantic entailment? What follows is pure speculation.

You might think of the triggering structure of the mind as layered, the attainment of some primitive concepts being the effect of (hence never prior to) the attainment of others. The facts of conceptual development in ontogeny and science would thus reflect—inter alia—the orderings imposed by this triggering structure. Since, pace the Empiricist, it would not reflect the order of logical construction, it would follow that there are logically possible organisms which discover (e.g.) their set theory before their arithmetic; only that we—given the nature of our psychology—are not among them. (Nor, it seems, are our children; vide the apparent failure of constructivist attempts to teach these subjects in reverse natural order.)

The idea is, then, that while all primitive concepts are ipso facto triggered, there is nevertheless a hierarchy of triggers, and it is this hierarchy, rather than the hierarchy of semantic entailments, which predicts the observed order of concept attainment. What’s inconvenient about this idea is, of course, that we can’t determine the structure of the triggering hierarchy in the way that we can determine the structure of the logical hierarchy—viz. a priori, by doing conceptual analysis. The hierarchical relations among triggers might, in principle, go any way at all, depending upon how the mind is put together. Whether there’s any practical use to the notion of a triggering hierarchy thus depends on whether we can find anything interesting to say about the interconceptual relations that it specifies; in particular, about which concepts are at which levels.

I think that we are, in fact, in a position to make some interesting guesses. I think we are in a position to argue like this: there are two, logically independent kinds of criteria which lowest (or, anyway, lower) level concepts in a triggering hierarchy ought to exhibit. If it turns out that there is a set of concepts which satisfy both these criteria (and if not all primitive concepts satisfy them), then that suggests there’s something of substance to be said about the structure of the hierarchy. And if it turns out that the concepts that are basic-level by these criteria are not the ones that are basic-level by the criterion of logical construction, then we have the groundwork for an empirical decision between the Empiricist and Nativist accounts of concept acquisition.

The two criteria which I think it’s reasonable to suppose that basic-level primitive concepts ought, in general, to satisfy are these: they should exhibit relatively high accessibility, both in ontogeny and in intellectual history; and they should be readily acquired by ostensive definition. I take it that the propriety of the first of these criteria is self-evident: Primitive concepts at the lower levels of a triggering hierarchy would be those whose acquisition requires relatively little mediation by other concepts previously attained. That is, the triggers for lower-level concepts should be largely experiential. So, given an appropriately rich environment, basic-level primitive concepts should be acquired early, and in no environment should they be acquired later than primitive concepts at higher levels.

The appropriateness of the second criterion may be less evident. I’m assuming, as a rough rule of thumb, that ostensive definability varies inversely with conceptual mediation: that the more easily a concept is ostensively introduced, the less the acquisition of that concept depends (causally) upon other concepts being acquired first. Quine has made something like this point, though in another context to be sure. He suggests that the notion observation sentence is to be reconstructed by appeal to the notion of ostensive definability, as follows: “the more observational a sentence is, the more nearly its use can be mastered ostensively” (1970, pp. 5–6). The idea is that the observation language is comprised of terms whose application is not mediated by theory; hence terms that can be learned without the mastery of the theoretical vocabulary; hence terms that can be learned by ostension.

This is a rule of thumb and should be taken with a grain of salt,
It seems to me that one of the most interesting facts currently available for constraining theories of concept acquisition is that these criteria do indeed pick out the same sorts of cases. In particular, as a number of psychologists have pointed out (see especially Brown, 1958; Rosch et al. 1976), there seems to be a natural ontogenetic bias in favor of concepts at what you might describe as intermediate levels of logical abstractness. Consider, for example, such implication hierarchies as:  Poodle → dog → mammal → animal → thing. If you suppose that lexical concepts are normally constructed out of their semantic entailments, then you might suppose that what children do is: they first learn THING, out of which they construct ANIMAL, out of which they construct MAMMAL, out of which they construct DOG, out of which they construct POODLE. Or, alternatively, they first learn POODLE, out of which they abstract DOG, out of which they abstract MAMMAL, out of which they abstract ANIMAL, out of which they abstract THING. The idea that concepts have internal structure can be (and has been) accommodated to either of these predicted orderings.

In fact what happens is neither. Rather, children tend to enter this hierarchy at the level of DOG and, as it were, work down to the subtended species and up to the subtending ones. Similarly, they get TABLE before they get either FURNITURE or END TABLE; they get CAR before they get either VEHICLE or VW; they get HOUSE before they get BUILDING or BUNGALOW, and so on across dozens and dozens of examples. Ontogenetic priority seems to go along, in these cases, with a cluster of other traits: the early attained concepts are often the most abstract members of their implication hierarchy which subtend objects of approximately similar shapes; they tend to be relatively high frequency, and, or so it seems to me, they tend to be especially readily acquired by ostension. For example, I take it that you could teach “dog” ostensively to a child who had not learned “animal” or “poodle”; but I bet you’d have a beastly time trying to teach “poodle” to a child who hadn’t acquired “dog”, or “animal” to a child who hadn’t prior mastery of some animal words at the same level of abstraction as “dog”.

It bears emphasis that this coincidence, rough and ready though it doubtless is, is really quite remarkable. There’s surely no logical
reason why HOUSE should be earlier acquired and more easily ostensively defined than BUILDING or BUNGALOW. Say, if you like, that the ease of ostensive definition explains the ontogenetic priority; but notice, if you do say this, that the explanation is quite vacuous: it leaves open the question why the concepts should differ in ostensive definability. The present proposal is that both facts are consequences of the structure of the triggering hierarchy and that the structure of the triggering hierarchy is a brute fact—i.e. that it's part of the aboriginal structure of the mind.

I want to draw two morals: the first is that the range of concepts that the criteria of ontogenetic availability and ostensive definability pick out as basic does not, so far as anybody knows, parallel the order of logical abstraction, and this fact is very hard to explain on the concepts-are-logical-constructs sort of story. If you take the Empiricist account really seriously, so that lexical concepts are typically (not just constructs but) constructs out of a sensory base, you make a travesty of the ontogenetic data. It just isn't true that (e.g.) color and shape concepts are learned early, or that they are easily ostensively defined. Most children are still having trouble with "red" long after they have COOKIE and DADDY and DOGGIE under their belts. Nor is this simply a problem about the acquisition of vocabulary, since young children have terrible trouble learning color-based discriminations in nonverbal versions of the concept learning task (see, for example, Carey 1978). And, even if you give up the idea that basic concepts are epistemologically specifiable, there isn't any way that I can see of predicting the order of ontogenetic accessibility from the facts of semantic entailment. What seems to be got first, and what seems to be readily acquired from instances, is, roughly, middle-size-object/event concepts of middle-level abstractness.

The second moral is that (pace Quine) the ordering established by the criteria of early attainment and ready ostensive definability isn't the ordering of grades of theoreticity either. Consider such folk-psychological concepts as ANGRY, SAD, HAPPY, etc. I think there's no doubt that these are acquired early, that they must have been part of the universal prehistory of our species, and that they are easily introduced by ostension. This isn't to say, however, that they are relatively peripheral in any ordering of the concepts by explanatory centrality. If you want to say how such concepts are related to overt behavior—what role they play in folk-explanations of what organisms observably do—you have to tell a very long story; one which analytical philosophers of mind are still trying to unravel. The vocabulary of folk psychology is theory vocabulary, but it's learned early and taught ostensively for all that.

The point is that the structure of the triggering hierarchy probably isn't natural from any philosophical point of view, though it may be quite natural from the point of view of an ethologist. If a social organism is to survive its infancy, it better have ready access to the concepts that structure its implicit theory of the mental states of conspecific adults; the triggering constraints on such concepts had better not require the prior acquisition of a rich conceptual repertoire. There is, then, a sense in which basic concepts are, ipso facto, close to experience: but it's not the sense of "logically constructed out of concepts which subextend experiences," and it's not the sense of "peripheral on the scale from observation to theory." Whichever of these ways you read it, the idea that, as it were, ontogeny recapitulates epistemology, is hopeless. On the contrary, the only sense in which basic concepts are close to experience is this: basic concepts are the ones that are normally triggered by encounters with their instances (whereas the triggering of nonbasic primitive concepts is mediated by the previously attained conceptual repertoire; and, of course, complex concepts—e.g. phrasal concepts—are learned by logical construction). If, therefore, you want to know which concepts are in fact normally triggered by encounters with their instances, you can't find out by doing conceptual analysis, and you can't find out by doing epistemology. You have to go and look.

"Some Rationalist you are," you may now want to say, "Why, you're nothing but a closet skeptic. For isn't your view really that the facts of theory construction and of concept attainment are precisely irrational? That the history of our science and the ontogeny of our concepts, is simply the reflexive manifestation of genetic structure under the impact of environmental stimulation? What reason do you have for supposing that our concepts apply—that our science is true—if warranted inference doesn't mediate their attainment?"
I have three answers to make. The first is that the truth of our science doesn't depend on how we came to discover it, only on its correspondence to the facts. The origin of a concept is, no doubt, a poor predictor of its warrant; so you should establish warrant by argument to the best explanation, not by committing the genetic fallacy. A Rationalist would predict that the history of theory construction, and of the ontogeny of our concepts, might well be a matter of sudden insights and flashes of intuition; of inspiration, in short. And that seems plausible enough, given the anecdotes. What clever people tell us about how they get good ideas is that ripeness is all: you expose yourself to the play of experience, and you keep your fingers crossed. In short, what clever people tell us is that new concepts do not emerge by induction and logical construction from the repertoire of concepts perviously attained. My point is that what clever people tell us makes perfectly good ethological sense and affects the truth of our science not one whit, so long as the traditional distinction between the logic of discovery and the logic of confirmation is scrupulously observed.

Nor need a Rationalist deny that the character of the concepts we entertain depends upon the character of the world we live in. No simple concept is available unless it is triggered, and while the function from triggers onto the concepts they release is innately specified, it is the way the world is that determines which triggers we in fact encounter. Our innate endowment determines which worlds we can, in principle, understand; but only the interaction of that endowment with the stimulations we receive determines which science we actually develop.

However, I do plead guilty to this: if the trigger-concept pairings are innately specified, and if the principles of construction are too, then the space of concepts that are available-in-principle is innately closed, and we have no guarantee that the concepts required to build a true science are situated in that space. But, though I have to admit this, so too—as far as I can tell—does everybody else. For example, if you think the space of concepts is the closure of the sensory concepts under the mechanisms of association, then you must hold that only worlds which can be described by sensory concepts and constructs out of them can, in principle, be the objects of our science. Indeed, I can't make much sense of the idea that our psychology could be so arranged that we would be able to develop the right science however the world might be. And, even if that idea does make sense, why on earth should anyone suppose that it is true? If that is skepticism, then I am quite prepared to live with it.

Everybody's a Rationalist in the long run. Everybody accepts the requirement for a primitive conceptual basis; everybody accepts the requirement for logical construction of infinitely many phrasal concepts; everybody accepts that primitive concepts are psychologically simple; everybody accepts that simply concepts are unlearned. What distinguishes Descarte's kind of Rationalism from, say, Locke's—insofar as they differ in their views about concept attainment—is an issue over how big the primitive basis actually is. Soi-disant Empiricists are inclined, in this respect, to niggle.

I see no interesting sense in which the size of the primitive basis is an a priori question—we immediately recognize it as ethological in the case of all other species—except insofar as the failure of a variety of reductionist programs in philosophy indicates that the primitive basis must be a lot bigger than was once supposed. Surely all basic concepts must be primitive, just as Locke and Hume maintained. Since epistemologically grounded programs of reduction always fail, it would seem to be reasonable to accept a cluster of non-epistemological criteria as prima facie indices of membership in the conceptual basis: ostensive definability, early acquisition in ontogeny and in intellectual history, lexicalization, and universality and relative independence from prior concept attainment are among these indices, and, very likely, research will turn up lots of others. Current evidence suggests that there must be thousands of concepts which satisfy these criteria: inter alia, concepts which subtend middle-sized objects and events. This population of basic concepts represents a very conservative estimate of the population of primitive concepts, since the former includes only the bottom level of the triggering hierarchy, whereas the latter includes all concepts that are psychologically simple. Since all primitive concepts must ipso facto be unlearned, it looks as though the innate structure of the mind is going to be very rich indeed according to the present proposal. Our ethology promises
to be quite interesting even if our developmental psychology turns out to be a little dull.

I don't expect you to be convinced. I am not convinced myself. For one thing, the evidence that nonsensory lexicalized concepts are typically unstructured would have to be much better than it now is before it would be rational to consider scrapping the Empiricist program in its entirety. I'll settle for this: to have convinced you that the issue about conceptual innateness is in the relevant sense empirical; to have convinced you that, in principle at least, relevant evidence can be drawn from a wide variety of research domains; to have convinced you that the question of the existence of logical and psychological structure in de facto lexical concepts is very close to the heart of the matter; and to have convinced you that, in the current state of our cognitive science, it is far from being an open-and-shut decision in favor of the concept learning model.

It seems to me that Anglo-American theorizing about concept attainment has, for several hundred years now, restricted itself to the consideration of a very small range of theoretical options. It also seems to me that the results have not been extraordinarily encouraging. Perhaps it is time to throw open our windows, kick over our traces, upset our applecarts and otherwise wantonly mix our metaphors. If we are going to have a cognitive science, we are going to have to learn to learn from our mistakes. When you keep putting questions to Nature and Nature keeps saying "no", it is not unreasonable to suppose that somewhere among the things you believe there is something that isn't true.

Notes

Introduction

1. Collins, in a recent (1979) paper, has called attention to the following consideration: though beliefs are causes and beliefs have truth values, still it is possible to doubt that there is anything which is both a cause and truth- valuable. The idea is that, in the sense of belief in which beliefs have effects, they are mental particulars; so that John's belief that P is numerically distinct from Jim's unless John = Jim. Whereas, in the sense in which beliefs have truth-values, they are abstract objects (e.g. propositions), so that if both John and Jim believe that it will rain, they share a belief; there is a belief that they have in common.

This distinction is both traditional and at least arguably correct, but it does very little to alter the general situation. Suppose that it is propositions that bear truth values. Then part of our theory of propositional attitudes will have to be an account of the relation that holds between me and a proposition when the proposition is one of the ones that I believe. So far as I can see, that relation will have to be mediated by a mental representation (viz. a mental particular to which causal properties are ascribed.) If that sort of treatment is correct, then the mental representation will, in turn, have the semantic property (not of being truth- valuable but) of expressing a proposition. So, on this model, there will after all be things that have both semantic properties and causal roles, and the question how anything with a causal role can have semantic properties will still have to be asked and answered. (It will have to be asked—and answered—anyhow, since token sentences—e.g. this token sentence—have both causal roles and semantic properties.) If you think this is an artifact of the model, you need some alternative way of construing the relation that holds between a proposition and someone who believes the proposition—a construal which, in particular, does not rely on the notion of mental representation. I know of none such that seems remotely plausible.
the formal semantics (type a), the account of the logical syntax of the vehicles of representation (type c), and the psychology of reference (type b) all fit together.

12. I wish to thank the many members of the AI community who read the manuscript and offered good advice (like “My God, you can't publish that!”). I'm especially indebted to Steven Isard, Philip Johnson-Laird, and Zenon Pylyshyn for illuminating comments on an earlier draft.

9. Methodological Solipsism

1. I shall speak of “type identity” (distinctness) of mental states to pick out the sense of “same mental state” in which, for example, John and Mary are in the same mental state if both believe that water flows. Correspondingly, I shall use the notion of “token identity” (distinctness) of mental state to pick out the sense of “same mental state” in which it's necessary that if x and y are in the same mental state, then x = y.

2. For extensive discussion, see Fodor (1975; and Chapter 7).

3. This is not, notice, the same as saying “formal operations are the ones that apply mechanically”; in this latter sense, formality means something like explicitness. There's no particular reason for using “formal” to mean both “syntactic” and “explicit,” though the ambiguity abounds in the literature.

4. Some fairly deep methodological issues in Artificial Intelligence are involved here. See Chapter 8 where this surface is lightly scratched.

5. I'm told by some of my friends that this paragraph could be read as suggesting that there are two kinds of beliefs: opaque ones and transparent ones. That is not, of course, the way that it is intended to be read. The idea is rather that there are two kinds of conditions that we can place on determinations that a pair of belief tokens count as tokens of the same belief type. According to one set of conditions (corresponding to transparent taxonomy) a belief that the Morning Star is such and such counts as the same belief as a belief that the Evening Star is such and such; whereas, according to the other set of conditions (corresponding to opaque taxonomy), it does not.

6. I'm leaving it open that it may be to say still less than this (e.g., because of problems about reference under false descriptions). For purposes of the present discussion, I don't need to run a line of the truth conditions for transparent propositional attitude ascriptions. Thank Heaven, since I do not have one.

7. It's worth emphasizing that the sense of “behavior” is proprietary, and that that's pretty much what you would expect. Not every true description of an act can be such that a theory of the mental causation of behavior will explain the act under that description. (In being rude to Darcy, Elizabeth is insulting the man whom she will eventually marry. A theory of the mental causation of her behavior might have access to the former description, but not, surely, to the latter.)

Many philosophers—especially since Wittgenstein—have emphasized the ways in which the description of behavior may depend upon its context, and it is a frequent charge against modern versions of Rational psychology that they typically ignore such characterizations. So they do, but so what? You can't have explanations of everything under every description, and it's a question for empirical determination which descriptions of behavior reveal its systematicity vis-à-vis its causes. The Rational psychologist is prepared to bet that—to put it very approximately—behavior will prove to be systematic under some of the descriptions under which it is intentional.

At a minimum, the present claim goes like this: there is a way of taxonomizing behaviors and a way of taxonomizing mental states such that, given these taxonomies, theories of the mental causation of behavior will be forthcoming. And that way of taxonomizing mental states construes them non-transparently.

8. One might try saying: what counts for opaque type individuation is what's in your head, not just what's running through it. So, for example, though Alfred and Misha are both thinking “he feels faint,” nevertheless different counterfactuals are true of them: Misha would cash his pronoun as: “he, Sam” whereas Alfred would cash his pronoun as: “he, Misha.” The problem would then be to decide which such counterfactuals are relevant since, if we count all of them, it's going to turn out that there are few, if any, cases of distinct organisms having type-identical thoughts.

I won't, in any event, pursue this proposal since it seems clear that it won't, in principle, cope with all the relevant cases. Two people would be having different thoughts when each is thinking “I'm ill” even if everything in their heads were the same.

9. I'm assuming, for convenience that all the Ps are such that either they or their denial are believed. This saves having to relativize to time (e.g., having (b) and (c) read “...you believe or come to believe...”).

10. I blush to admit that I had missed some of these complexities until Sylvain Bromberger kindly rubbed my nose in them.

11. It may be that Putnam does accept this moral. For example, the upshot of the discussion around p. 153 of his article appears to be that a Greek semanticist prior to Archimedes could not (in practice) have given a correct account of what (the Greek equivalent of) “gold” means; viz., because the theory needed to specify the extension of the term was simply not available. Presumably we are in that situation vis-à-vis the objects of many of our thoughts and the meanings of many of our terms; and presumably, we will continue to be so into the indefinite future. But, then, what's the point of defining psychology (semantics) so that there can't be any?

IV. Nativism

10. The Present Status of the Innateness Controversy

1. A concept is an abstract entity on anybody's view in that concepts are types. The interesting question is whether they are types whose tokens are mental objects. Frege seems to argue—in “The Thought” (1968)—that con-
cepts are abstract, hence not mental. I can't imagine why he found this line persuasive.

2. Strictly speaking, what with concepts being types, one ought to say that primitive concepts arise from the transducer outputs via a process of "abstraction". I shall usually ignore this, since it complicates the exposition without much clarifying the issues. Suffice it that such abstractive processes are not to be identified with the combinatorial processes lately discussed. Combinatorial processes take simple concepts onto complex ones, whereas abstraction takes transducer outputs (what used to be called "experiences") onto concepts in the primitive basis and, more generally, tokens onto the types that they are tokens of.

3. According to most traditional Empiricist accounts, there are two ways in which a complex concept can be acquired other than from experience: by explicit (verbal) definition, and by the operation of the faculty of "imagination"; which functions to produce complex concepts by freely combining primitive ones. (The concept UNICORN is often cited, but any concept that is both complex and uninstantiated would do as well.) None of this bears on our main line of inquiry, however, and I mention it largely in a spirit of comprehensiveness.

4. This way of putting it takes the relevant hypotheses to be ones that express identity conditions on a concept: the hypothesis refers to a concept and says what its analysis is (the concept flurg is the concept of something which is such and such). It's more in the spirit of some Empiricist writing (e.g. of Locke's analysis of "framing" a concept) to think of concept attainment as simply the bringing together (e.g. by association) of the constituents that belong to the analysis of the concept. So, suppose the constituents of the concept A are B, C, and D. Framing the concept A would then be a matter of forming an association between B, C, and D. The point is that the formation and confirmation of analytic hypotheses still come in, though the content of the hypotheses is relevantly different. In the present case, to frame the concept would be to confirm something like: things that satisfy B and C typically satisfy D. There are, I think, reasons why an Empiricist is well advised to set things up the way I have them in the text, but the issue is, in any event, neutral on the questions that distinguish between the Nativist and the Empiricist view. (I am indebted to Charles Marks for bringing this point to my attention.)

5. I say this is plausible, but I don't guaranty that it's true. What the claim amounts to is that, for purposes of such inductive processes as are involved in concept acquisition, the data base can be specified solely by atomic sentences in the primitive vocabulary. I think that it's characteristic of Empiricists to assume this tacitly, both in psychology and in confirmation theory. The issue occasionally becomes explicit, as in the discussion of "negative facts": If there are negative facts, then presumably some observation sentences are molecular (they have the form (not P)).

6. I put the point in this uncharacteristically cautious way because there have been a variety of Rationalist theorists—e.g. Plato, Descartes, Katz—who apparently want to appeal to the internal structure of concepts as part of their account of conceptual necessity, but who nevertheless do not hold that complex concepts are learned by processes involving logical construction. Indeed, I think that many philosophers are convinced that concepts must have internal structure precisely because they are convinced that some truths are conceptually necessary. It is, however, increasingly clear that the argument that goes: "bachelors are unmarried" is necessary, therefore UNMARRIED must be a constituent of BACHELOR" is simply a non sequitur. (See Fodor, Garrett, Walker and Parkes 1980, where this issue is discussed at length.)

7. By the way, the primary evidence that most phrasal concepts are complex is that most phrases are definable: "is a brown cow" = df "is brown and is a cow"; this consideration would lead us to treat phrasal concepts as typically complex even if we didn't have the productivity problem to cope with. The application of the reduction program to phrases has, in general, gone brilliantly; whereas its application to the lexicon has, in general, gone nowhere. That, too, should worry an Empiricist.

8. What this formula doesn't do, of course, is provide criteria for telling whether something is a tiger. I think that much of the argument in "The meaning of meaning" rests on an equivocation between these two senses of "determine the extension". What Putnam's arguments seem to me to show is that the internal representation of a term needn't provide criteria for its application; they are, in fact, anti-verificationist arguments. But what Putnam infers from his demonstration is that the internal representation of a term needn't be, as it were, coextensive with that term—e.g. that something might be a tiger (so that "tiger" is true of it) and yet be such that the concept TIGER doesn't apply to it. This latter conclusion certainly doesn't follow from merely anti-verificationist premises, nor can I think of any other arguments that would lead one to embrace it.

9. Empiricists, too, accept that the coincidence of (a-d) defines a level of the triggering hierarchy, since sensory concepts satisfy all four conditions. Viewed from this perspective, the difference between Empiricist and Rationalist accounts is that the former postulate a degenerate triggering hierarchy; one in which the level of basic concepts is the only level there is.