ROTE LEARNING AND LEARNING WITH UNDERSTANDING

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF PHILOSOPHY

AND THE COMMITTEE ON THE GRADUATE DIVISION

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Hugh Gilbert Petrie

August 1965

WL

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John I. Matthews Jr.

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Patrick Suppes

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ernest R. Hilgard

(Psychology)

Approved for the University Committee
on the Graduate Division:

Virgil K. Whitaker

Dean of the Graduate Division

ACKNOWLEDGMENTS

I want to express my appreciation to several people who have helped me enormously with this dissertation. First, the members of my committee, John Mothershead, Patrick Suppes, and Ernest Hilgard have each, in his own way, given me much advice, encouragement, and criticism. I could not have succeeded without them. I also want to mention John Goheen, Lawrence Thomas, and Israel Scheffler whose thoughts have helped inspire the general direction of this work. And last, but above all, I want to acknowledge the encouragement, patience, understanding, and wisdom of my wife, Patricia. It is to her and our children that this work is dedicated.

TABLE OF CONTENTS

CHAPTER I

A CONCEPTUAL ANALYSIS OF ROTE LEARNING AND LEARNING WITH UNDERSTANDING

(1) Max has learned to recite the multiplication tables by heart, but he certainly hasn't learned much about arithmetic. On the other hand, Sally has learned what arithmetic is all about.

(2) Judy has learned to recite the text-book answer that slavery was an issue in the Civil War but she doesn't understand it. Saul, however, has not only learned that slavery was an issue in the Civil War, but also why it was and how it was related to other factors. He really understands the Civil War.

We commonly make statements like (1) and (2) in talking about learning. These statements are usually meant to characterize a difference between merely learning something by heart and learning it with understanding. In this chapter I shall attempt to explicate the difference between these two notions and to see how they are related to other concepts such as knowledge, belief, reasons, intention, practice, etc.

1. Some Possible Attempts at Explaining the Difference between Rote Learning and Learning with Understanding

The first question that arises is whether this apparent distinction between rote learning and learning with understanding is one of degree or of kind. In other words, is the distinction to be understood in terms of a single meaning of 'learning' with the difference drawn in terms of how this learning is accomplished or retained or exercised, or is the distinction rather due to a difference in the logical relations of the two learnings, i.e. to an ambiguity in the meaning of 'learning'?

A first attempt at explaining the apparent difference in terms of a univocal meaning of 'learning' is the position which says in effect that (1) and (2) simply show differences of degree and not of kind. Thus, according to this line, (1) might be roughly paraphrased as

> Max has only learned to recite the multiplication tables so far. He might learn what multiplication is all about later like Sally, or he might not, but in any case learning about multiplication in general is, if you will, just like learning to recite the multiplication tables, only it is more complex and anyone who has learned it can do many more things as well. For instance, Sally can not only recite the multiplication tables, she can use them in solving practical problems and she can also explain the relation of multiplication to addition and division.

(2) might become

> Saul's learning is quite detailed. He not only can answer "Yes" to the question 'Was slavery an issue in the Civil War?', he can also answer many other questions concerning slavery, the Civil War, and

their relations. Judy can only give the text-book answer that slavery was an issue in the Civil War, but not Saul. He has learned this and much more besides.

The central line behind these answers seems to be that the only difference between "rote learning" and "learning with understanding" is that the latter contains quite a number of bits of learning while the former contains only a few. There is a difference of degree (the number of bits of learning involved) but not one of kind. One already begins to feel a bit uneasy about characterizing "learning with understanding" in this way, since we ordinarily think we mean something more by it than the mere ability to recite large numbers of statements, but perhaps the story could be amplified to make it sound more plausible.

One way of doing this might be to use the vocabulary of set theory and stimulus-response psychology. If we take learning to be identified with a change in the class of responses available to a person given some stimulus, then the difference between rote learning and learning with understanding might be explained in terms of the mere size of these classes. Rote learning is identified by a very small response class, while learning with understanding has a much larger class. Now, although this procedure seems somewhat artificial (after all, what criteria do we use in judging a class to have a "large enough" number of responses), nevertheless, such a procedure does seem to

capture at least one of our intuitions concerning rote learning versus learning with understanding. This intuition is that rote learning is of somewhat limited value and of use in only very narrowly circumscribed situations. On the other hand, learning with understanding is intuitively thought to be much more valuable through being of wider application. However, attempting to capture this intuition by the mere size of the response class seems to suffer from the intuitive objection that it is not so much the size of the response class, as it is the quality of what is in it.

For example, the actual size of Max's response class using the multiplication tables may be very large. He might be able to recite the tables clear up to 100 x 100 and apply them in many practical situations. On the other hand, Sally's response class may be very small. She might not know any more than up to 9 x 9, but if her response class also contains such statements as express the theoretical relations of addition and division to multiplication, the notion of natural numbers having a successor, etc., few would hesitate, I think, in saying that she understood arithmetic better than Max despite the supposed near equality of size of response classes.

A similar situation would occur with my second example. Judy may have memorized the whole text book while Saul knows only a few general propositions. But if Saul's knowledge is of a key nature and he can use it in explaining

the Civil War, then surely he too has learned with understanding.

Another approach in this vein might attempt to tie the difference between rote learning and learning with understanding to whether or not the learning requires practice.[1] Thus a reference to rote learning would implicitly carry with it a reference to having practiced the thing learned while learning with understanding would not. In either case reference would be made to two different things, but the difference would be an empirical matter and the meaning of 'learning' would in either case be the same.

Thus to recur once more to my example, Max had to practice very hard in order to learn to recite the multiplication tables. His present ability to do so is simply a result of rote learning. He also has not learned with understanding what multiplication is all about nor how it fits in with other arithmetical conceptions and principles. But if he did learn these further things, as Sally has done, he would not have to practice them. There would simply come a time when he "saw" that division is the inverse operation of multiplication and he would not have to practice this "seeing" at all. The concept of practice seems

---

[1] See J. Roland, "On Knowing How and Knowing That," Philosophical Review, Vol. 67 (1958), 379-388, for an application of the "practice" criterion to distinguish two sorts of "knowing how." It will be obvious that this is closely connected to what I am considering here.

to be wholly inadmissible in this case. Even if the inverse concept only slowly dawned upon him over a period of time, we would not, I think, be willing to say that he was practicing anything during this time. Sally, on the other hand, did not require "practice," in any obvious sense, to understand arithmetic. She simply read the lesson, perhaps thought about it awhile, and "saw" what was going on. Again the situation is entirely similar with the case of Judy, Saul, and the Civil War.

The "practice" formulation also seems to capture a part of our intuitions concerning rote learning and learning with understanding, for most pieces of learning which we want to call "rote" are of the nature of "skills" which require a certain amount of practice to get right. On the other hand, learning with understanding, although it often requires a great deal of thought and study, intuitively seems not to require anything that can be called "practice." However, the "practice" formulation faces several difficulties. I shall mention two of these. First of all, there are some skills, e.g. doing logic and playing chess, which, while they require practice, also involve, at least intuitively, a great deal of learning with understanding as well. On the other hand, some simple skills seem to require very little, if any, practice. It seems entirely possible that one could learn to tie a particular knot simply by seeing it done once. The second difficulty concerns the notion

of "practice" itself. We do not seem to want to call the study and thinking which may precede learning with understanding "practice," but why is this? The notion of "practice" would have to be spelled out in some detail before it could reasonably serve as a distinction between rote learning and learning with understanding.

A very promising notion which is not unrelated to the previously discussed notion of response classes is that of "transfer," as used in psychological writing on learning. However, the concept of transfer seems not to have any precise logical characterization. Intuitively a piece of learning has a large transfer value if it can be applied to new situations; situations rather unlike those in which the learning originally took place. That phenomena exhibiting this intuitive notion do take place seems not to be contested by psychologists. However, the specification of transfer seems to be not very well agreed upon at all. "Transfer" has any number of very precise senses, given by the various methods of measuring it proposed by psychologists. For each psychological experiment using the concept, there seems to be a precise characterization of what the experimenter takes transfer to be in the <u>particular</u> experiment. Two of the more popular characterizations are the following: One might measure the time required to learn two tasks separately by two control

groups and then the time required to learn them successively by a third and fourth group, A before B and B before A. Any savings by the successive learning is transfer. Another approach might involve learning how to do algebra by two methods. Then a test which does not use directly any of the specific problems or techniques learned in either method might test the relative transfer value of the two methods for the problems on the test. There are other methods as well, but the philosophical problem remains untouched--that is, the more exact specification of just what our intuitive notions of "transfer" mean in general, so that a clear picture can be gotten of how these various precise empirical tests do measure transfer, if at all.

What is important for my purpose about the notion of transfer is that it seems to capture another intuitive notion we have about rote learning and learning with understanding. Those bits of learning which have a high transfer value we tend to call learning with understanding, while those bits having a low transfer value are generally put in the rote learning category. Unfortunately, lacking a clear meaning of 'transfer' it is not at all obvious whether transfer marks a difference in the kind of learning or merely a difference of degree.

For example, Hilgard says,

> Because all learning is to some extent cognitively controlled, the distinction between blind learning

> and learning with understanding becomes one of degree. There is a point, however, at which understanding takes on new prominence. That is in the kind of creative attack on a problem in which experience does not provide solutions ready-made, but the solutions have to be discovered or invented through a reconstruction of experience appropriate to the problem at hand. This is not completely specified by experiments on insight, but it is suggested by them.[2]

Is the point "at which understanding takes on new prominence" to mark a difference in the meaning of 'learning' in rote learning and learning with understanding, or is it, as Hilgard seems to imply, only a difference in degree? I think that this point does mark an ambiguity in 'learning' and my reasons for this will become evident in the following. However, for now, let me suppose that transfer really does only mark a difference of degree. We might suppose that what is learned has a high transfer value and thus is learning with understanding, or a low transfer value and is rote learning. Further, the presence or absence of transfer does not mark a difference in the meaning of 'learning' since both are learned, simpliciter.

To recur for the last time to my examples we can see that Max's learning of the multiplication tables has low transfer value, although presumably it does have some, i.e. I am supposing that he can utilize his learning in figuring out how much 9 apples at 8¢ each will cost,

---

[2]E. R. Hilgard, Theories of Learning (1st ed., New York: Appleton-Century-Crofts, 1948), p. 343.

although it is certainly possible (both logically and empirically) that his memorization of the multiplication tables extends only to answering correctly what 9 x 8 is and not to any practical application. Sally's learning, on the other hand, has high transfer value and is thus learning with understanding. She can talk about multiplication in a variety of contexts. Saul's learning has very high transfer value in that he can discuss a number of questions about the Civil War and very probably can make some highly interesting and instructive generalizations about the causes of war in general. His learning is learning with understanding, but Judy's is not. She can only give the text-book answers.

2. Another Approach toward Explaining the Difference between Rote Learning and Learning with Understanding

To begin with, all the approaches which I considered in the last section, although by no means all the possible ones, seemed to have two things in common. First, they tended to treat the distinction between rote learning and learning with understanding as exhaustive. All learning was either with understanding or by rote. Second, with the possible exception of the transfer approach, they all seemed to assume that 'learning' was wholly ambiguous and any difference between rote learning and learning with understanding was to be a difference in degree of <u>something</u>

or other and not a difference of kind.

In this section I want to question both of these assumptions. I shall try to develop an argument for claiming that the analysis of rote learning is logically distinct from that of learning with understanding. In the course of the argument, it will become apparent that this dichotomy will not serve to classify all learning exhaustively. There are numerous other classes as well and my analyses should serve only to throw light on some of the important features of the more common of these.

The method I shall use will be to isolate some "paradigm locutions" involving rote learning and learning with understanding and try to analyze these, keeping in mind the intuitions which the various approaches of the preceding section seemed to capture. By "paradigm locution" I mean nothing more than those forms of words which seem in ordinary language peculiarly suited to express the concept involved. I would also claim that if a paradigm locution can be found expressing, say, learning with understanding, then most other ways of talking about learning with understanding can be paraphrased into some form of the paradigm locution with very little, if any, loss of meaning. Furthermore, the analysis of the paradigm locution will then enable me to set forth clearly the logical features of the concept in question.

Obviously this approach is a sort of "ordinary language" approach in that it does seek clues to the analyses of the concepts in the way these concepts are expressed in ordinary language. However, I want to make it plain that I do not intend my analyses ultimately to rest upon an analysis of ordinary language as some Oxford philosophers seem to make their analyses rest on "what the ordinary man would say." Rather I use ordinary language as a <u>clue</u> only to what the important logical features of a concept might be.

Following this approach I want to ask just what a paradigm locution for learning with understanding might be? There seem to be at least four fairly commonly used learning locutions--'learning that', 'learning why', 'learning to', and 'learning how'. A moment's reflection should convince one that the latter two find much of their application in talking about learning simple responses and skills of varying complexity. Thus one learns how to type and how to swim. The paradigm uses of 'learning how' seem to fit most naturally the mold of performing a series of actions without being able necessarily to give reasons for such performances. Skills like typing are more commonly used after 'learning how' than are ones like logic. The case is similar with 'learning to' and I shall postpone discussion of these two until I come to the analysis of rote learning which they seem to fit better.

However, both 'learning why' and 'learning that' are often used in reporting learning with understanding. Thus we say, "He has learned why there are two hydrogen atoms in each molecule of water," and "She has learned that naive set theory is inconsistent." Of these two locutions I shall take 'learning that' as my paradigm for analyzing learning with understanding. There are two reasons for this. First, it seems that most 'learningswhy' are summary reports of more specific 'learningsthat'. Thus, "He has learned why there are two hydrogen atoms in each molecule of water <u>because</u> he has learned that hydrogen has a valence of plus one and oxygen a valence of minus two. Furthermore, he has learned that the chemical valence must be zero in any stable chemical compound." It seems that the 'learning that' locution expresses the <u>specific</u> reasons justifying the use of the 'learning why' locution.

To put the point in a slightly different way, when we are told that someone has learned <u>why</u> something is the case, it seems we can always ask what it is <u>that</u> he has learned which gives the explanation implicitly guaranteed by the 'why'. On the other hand, when we are told that someone has learned <u>that</u> something is the case, we do not seem to be able to ask so easily what it is that he has learned <u>why</u> which justifies the 'learning that'. We can, of course, ask him why he believes that which he has

learned, but this seems to be asking for other 'learnings that' which justify the original one.

Second, using the 'learning that' locution enables me to compare my analyses quite easily with recent analyses of other obviously related notions such as 'knowing that', 'believing that', and 'teaching that'.[3] I shall therefore take 'X has learned with understanding that p' as my paradigm locution for analyzing learning with understanding.

A few remarks on 'learning that', simpliciter, will be in order first. It seems obvious that if someone has learned that water is a compound of hydrogen and oxygen, then he also believes that water is a compound of hydrogen and oxygen. That is, I claim that 'learning that' entails 'believing that'. To see this, let me consider two examples, the first where the person does not believe what has ostensibly been learned, perhaps because he reserves judgment, and the second in which the person actually believes the contradictory of what has ostensibly been learned.

Suppose that John reads a book or article by George Gamov on the origin of the universe. (Gamov holds that the universe originated with a "big bang.") Now we might easily say that John has learned that the universe originated

---

[3]See, for example, Gilbert Ryle, The Concept of Mind (New York: Barnes and Noble, 1949), for a classic comparison of 'knowing that' and knowing how'.

with a huge explosion. However, further suppose that John is also aware of the rival "steady state" hypothesis concerning the origin of the universe. (Very simply, this hypothesis admits that the universe is expanding but claims that it has always been expanding.) Being a careful man, John therefore simply withholds <u>any</u> belief whatsoever regarding the origin of the universe. It would seem that we here have a case in which John has learned that the universe originated with a huge explosion but he does not believe that the universe originated with a huge explosion because he has simply reserved judgment. However, appearances are deceptive, and what I want to claim is that John has <u>not</u> learned that the universe originated with a huge explosion but has learned that George Gamov, among others, holds a theory which <u>says</u> that the universe originated with a huge explosion. In other words, my original description of John's learning was really elliptical for this last description. When someone learns that such and such is the case, but does not believe it, he really learns that someone or some theory <u>says</u> that such and such is the case.

But now suppose that John actually believes that the steady state theory is correct. How should we describe his learning in this case when he reads Gamov's book? Again, I think the answer provided above is obviously correct. John learns that Gamov <u>says</u> the universe originated

with a huge explosion. There is, however, an interesting variation on this case. Suppose that although John believes that the steady state theory is correct, he reads Gamov's book in a course taught by a pedantic disciple of Gamov who is so benighted as to require that one answer test questions in accordance with his master's teachings. Thus if someone read one of John's tests in this course, wouldn't he be justified in saying that John has learned that the universe originated with a huge explosion? After all, John said so on his test. I think not. In this case, John has not only learned that Gamov <u>says</u> that the universe originated with a huge explosion, he has also learned that his teacher will not give him a passing grade in the course unless he (John) says so on his tests. Thus he also learns <u>to</u> answer test questions from this teacher with the desired answers.

This last bit of learning, 'learning to', will be examined in more detail when I analyze rote learning. At any rate I think that my line of argument is clear. In any case which might seem to contradict the entailment of 'believing that' by 'learning that' I will claim that the 'learning that' locution originally used is elliptical for 'learning that someone <u>says</u> that' or something similar, plus, perhaps, some additional learning. Thus I claim that 'X believes that p' is a necessary condition of 'X has learned that p', and therefore, also of 'X has learned with

understanding that p'.

I want to turn now to a consideration of the following two sentences:

(a) Paul has learned that the earth is flat.

(b) Sue has learned that the earth is spherical.

What I am interested in is whether the truth or falsity of 'p' in 'X has learned that p' makes any difference to whether 'p' can be learned with understanding or not. At first glance it would appear that the truth or falsity of what is learned makes no difference at all. Paul has learned that the earth is flat, believes it, makes it a part of his cognitive repertoire, and occasionally acts on it. This would be quite clear if Paul lived before Columbus' time. On the other hand Sue has learned that the earth is spherical, she believes it, has made it a part of her cognitive repertoire, and occasionally acts on it. Sue has learned something which is true while Paul has not, but surely this makes no difference in the learning that is involved.

Now while I think that such a characterization may be correct in some cases, I do not think that it is right in the case where the learning is learning with understanding. To see this let me examine a particular body of knowledge, say, elementary astronomy. It seems that there is some set of statements expressing the truths of

astronomy. Now I admit that this class is rather ill-defined, but that will not matter for my purposes. It also seems that for it to be true to say that someone has learned some astronomy, the person must have learned *that* the members in a certain subset of these sentences are the case. The person may have to learn certain other things as well, but this is irrelevant here. (I shall discuss the learning of a subject matter in more detail in Chapter III.) Again the subset of these facts which must be learned in order to have learned astronomy is not unique. That is, one could imagine one person learning one subset and another learning a different one and yet both of them learning astronomy. I would even admit that the two subsets might be mutually exclusive. But it nevertheless seems to be the case that *some* subset must be learned in order for us to say that the person has learned astronomy, and *a fortiori* for us to say that he has learned it with understanding.

It also seems quite clear that the person may learn some false things about astronomy as well as a great number of true ones. But if we are to say that the person has learned astronomy, the set of statements learned cannot contain very many false ones. Furthermore, it intuitively seems that the more false statements learned, the less the understanding. What I am saying is that although there may be no unique set of *basic* facts which must be learned

in order to learn astronomy, if someone has learned astronomy, then there is a set of facts which has been learned and most of these facts must be true. If this is so, then it seems reasonable to require the truth of what is learned as a necessary condition of 'X has learned with understanding that p'.

Several comments must be made. First, I distinguish between learning a particular fact with understanding, e.g. that the world is spherical, from learning a subject matter with understanding. It is the former to which the truth condition is to be applied and not the latter. The way in which the truth of the statements of a subject matter enters into understanding that subject matter will be discussed in the following. Second, I am aware that the justification I have given so far for the truth condition is rather weak. One might easily object to this requirement by saying,

> But surely one could learn Ptolemaic astronomy with understanding although it is false. In fact, keeping in mind your distinction between learning a particular fact with understanding and learning a subject matter with understanding, it is even the case that one could learn with understanding that the sun revolves around the earth.

I think this last sentence sounds a little odd, but I do admit that I must take care of objections like this, and I believe that I can. However, I must first develop my analyses more fully, so I shall for now simply continue treating the truth of what is learned as a necessary

condition of learning with understanding and take care of this sort of objection in the next chapter. At any rate, truth is quite obviously irrelevant to rote learning. What is rote learned may be either true or false.

I have so far claimed that necessary conditions for 'X has learned with understanding that p' are 'X believes that p' and 'p' is true'. Let me now turn to a third condition. It seems intuitively clear that learning with understanding must also include reasons for the belief held. It is not enough to understand something merely that one has come to believe it, even if that belief is true. To use Plato's famous example, a jury member may believe that a defendant is innocent and it may be the case that the defendant is innocent, but the jury member has come to have this belief solely through the eloquence of the defendant's lawyer. He has no justification for his belief, true though it may be. He has not learned with understanding that the defendant is innocent. For these reasons I propose 'X has good reasons for believing that p' as a third necessary condition of 'X has learned with understanding that p'. My analysis now looks like this: If X has learned with understanding that p, then

(1) X believes that p

(2) 'p' is true

(3) X has good reasons for believing that p

But (1)--(3) are just the conditions of an analysis of 'X knows that p' which goes back to Plato and which has been accepted by almost every philosopher since.[4] Thus a necessary condition of someone's learning something with understanding is that he comes to know it. This may not be surprising, but it enables me to use much of the recent work on 'knowing that' in my analyses. This identification of these three conditions of 'learning with understanding that' with the classical analysis of 'knowing that' does have some disadvantages too. For one thing there is the whole question of "non-inferential" knowledge. It is not altogether clear that 'knowledge that' must be analyzed in terms of the having of reasons. I shall discuss this point in more detail in the following chapter, but it does seem clear that if any knowledge without good reasons is possible, it is not of the kind which can be learned with understanding, for "understanding" clearly involves a direct reference to the possession of reasons of some sort. A second possible disadvantage centers around the existence of examples like "John knows that he is married." It seems quite conceivable that John could truly believe he is married with good reasons, and yet, _a priori_, it would seem odd to say that

---

[4]Plato, Theaetetus. See A. J. Ayer, The Problem of Knowledge (New York: MacMillan, 1956), R. M. Chisholm, Perceiving: A Philosophical Study (Ithaca, N.Y.: Cornell University Press, 1957), and Jaako Hintikka, Knowledge and Belief (Ithaca, N.Y.: Cornell University Press, 1957), for recent statements of this analysis.

he has learned that he is married. These two points seem to indicate that my conditions may be necessary but not sufficient for 'learning with understanding that'. Although I shall give a more detailed discussion of these problems in the next chapter, it seems to me that whether I can precisely formulate sufficient conditions is not too crucial, for my inability to do so would rest on odd sorts of problems such as these. And although it may be of philosophical importance to consider knowledge which is not "learned," nevertheless, for most ordinary cases of knowledge we know perfectly well whether they are likely to have been learned or not.

3. Some Possible Objections Answered and Easily Misconstrued Points Clarified

Before proceeding to a more detailed examination of my analysis, I would like to try to clear up some possibly troublesome points. Upon a superficial reading one might be tempted to say that what has been learned can be forgotten and thus the mere fact that someone has learned something is no guarantee that he now knows it. However, this is simply to ignore the fact that English almost always uses the present perfect to refer to actions in the past whose effects have continued up to the present. It is in this sense that I am using '<u>has learned</u>'. I take this to mean that what has been learned has not yet been forgotten,

and at any rate, I shall stipulate that this is the meaning I attach to my use of the present perfect, since it is by far the most common one. I shall use the past perfect to refer to occasions when forgetting has occurred. Thus "Susan had learned that such and such is the case" is most commonly followed by some statement to the effect that "however, she has now forgotten it."

A second possible criticism might run as follows:

> But certainly we often say that someone has learned that something is the case and yet we would not be willing to ascribe knowledge to him at all. For example, it might easily happen that near the beginning of a course on the Civil War, Jim learns that slavery was an issue in the Civil War. Just after such learning we could say that Jim has learned that slavery was an issue in the Civil War, and yet we might be loath to say that Jim knows that slavery was an issue in the Civil War. He might or he might not know it.

It is clear that even to give such a supposed counter-example any plausibility at all, one must use such phrases as 'near the beginning' and 'shortly after such learning'. It is the use of such phrases that give the whole show away, for these phrases clearly indicate that the learning which has been going on is not learning with understanding. It must be recalled that I have pointed out that 'learning with understanding that' typically occurs in the context of having learned a discipline or a subject matter and clearly the proposed counterexample is not such a case. The subject is only being begun, and the counterexample

loses all its force when it is told of Jim at the end of the course having learned some American History. In that case I do not see why we should not say that Jim knows that slavery was an issue in the Civil War.

Next, let me say something about the three conditions themselves. Much of Chapter II will be devoted to analyzing these conditions in some detail, but a few preliminary remarks would seem to be in order here. To begin with, the "good reasons" clause is especially in need of further analysis, as the explication of what it is to be a "good reason" is perhaps the major dividing point of a great many epistemologies. However, in my analyses I do not want to become involved in epistemological quarrels concerning the form which "good reasons" must ultimately take. I hope to remain neutral on this point. For example, it will not matter to my account whether the good reasons are propositional in form. They often will be (e.g. I take such things as statements by authorities in certain situations to be good reasons), but as to whether these good reasons are ultimately propositional or instead the having of sense data, whether they must have an element of certainty about them or not, these are matters which will either be immaterial to my account or will be treated later.

Concerning the "truth" condition, I simply follow Tarski's semantical conception of truth with neither apology

nor justification.[5] This conception seems so eminently plausible that at the very least it deserves to be carried through in various related analyses to see the range and power of its applicability without worrying about any of the controversies which may still surround it. For example, I do not want to get involved in the controversy concerning the consistency or inconsistency of ordinary language, nor the problem of how to formulate Tarski's admittedly meta-linguistic notion of truth in ordinary English. I do not think that such formal questions need concern me here.[6]

However, I do want to say a word or two about why I have separated the truth condition from the good reasons condition. It seems that in some cases of knowledge the reasons that a person has for believing a proposition are sufficient to demonstrate its truth, e.g. the reasons may be the steps in a valid sound deductive argument leading to the proposition in question as conclusion. In such a case the person not only knows the proposition, he knows that he knows it. On the other hand such cases are relatively rare. Most of the reasons which justify what we claim to know are

---

[5]A. Tarski, "The Semantic Conception of Truth and the Foundations of Semantics," Philosophy and Phenomenological Research, Vol. 4-5 (1944), 341-376, and "Der Wahrheitsbegriff in den Formalisierten Sprachen" (German trans. of a book in Polish, 1933), Studia Philosophica, Vol. I (1935), 261-405.

[6]See A. B. Levison, "Logic, Language, and Consistency in Tarski's Theory of Truth," Philosophy and Phenomenological Research, Vol XXV (1965), 384-392, for a discussion of how Tarski's theory of truth can serve to capture some of our ordinary notions of "true in English."

not "conclusive" in this sense at all. Most empirical propositions are of this problematic kind. Surely it is very unintuitive to say that we never know a proposition unless it has been conclusively demonstrated. Our store of knowledge would be miniscule if this were required. However, there is a sense of "conclusive" which simply means "to conclude"--to put beyond a reasonable doubt, not beyond a logical doubt--and in this sense, one could have conclusive reasons for believing a proposition which may, nevertheless, turn out to be false. The reasons are simply such as to make further present inquiry pointless, although in the face of new evidence, the inquiry could, of course, be reopened. It is this sort of good reasons which I have used as a condition for learning with understanding and knowledge, and it is clear that these reasons need not guarantee the truth of that for which they are reasons.

Yet it must be obvious that we wish to retain the truth condition for knowledge; for if someone believes something with good reasons and it later turns out to be false, we want to be able to say that he didn't really know it after all. It should be noted in this connection that I have offered an analysis of knowledge and not of the situation in which one is justified in claiming to know. It seems clear enough that someone may be justified in claiming knowledge without actually knowing. I shall

examine this distinction in some detail in the next chapter, and try to show there that most attempts to eliminate the truth condition from knowledge confuse "knowledge" with "being justified in claiming knowledge."

As far as further analysis of the "belief" condition is concerned, 'believes that' will be one of my primitives. I shall use it as a generic term to include a whole host of perhaps more psychologically explicit predicates such as 'fears that', 'hopes that', 'suspects that', 'is certain (in the sense of having a psychological feeling of certainty) that', 'half-knows that', 'feels that', etc. It should not be thought, however, that I am merely stipulating a definition of 'believes that'. It seems clear that all of the locutions I have listed as possibly more explicit formulations of 'believes that' have this in common--their cognitive content usually can be expressed as well, though perhaps somewhat less informatively, by the 'believes that' locution.

I am also aware that 'knows that' is sometimes used, especially by psychologists, to mean the same as what I would prefer to express by the locutions 'believes that with a feeling of certainty', or 'is certain (feeling of certainty) that'. Thus in this use, one could say "John knows that babies come from storks (usually with a firm tone of voice and an emphasis on 'knows'), but, of course,

they really don't." It might be thought that such a use proves to be a counterexample to my proposed analysis of 'knowing that', since what is "known" in this case need not be true. That it is not a counterexample is easy to see. Even the usual linguistic emphasis on "knows" in stating the case is sufficient to show that 'knows' is here being used in an inverted commas sense. "John doesn't <u>really</u> know that babies come from storks, he only thinks (believes) he does," is the right response to make, and not something like "He knows it all right, but it just isn't so."

However, in order to make my account complete, I must explain how such an inverted commas use of 'knows' comes about. The explanation is not far to seek. According to the analysis of 'knowing that' which I have given, that which is known is not only believed with justification, it is also true. Now suppose someone learns something with understanding and hence knows it. As this true proposition is used in guiding conduct, in deriving other true propositions, in short, in working with this piece of knowledge and in using it in our lives, we often begin to have feelings of certainty associated with it, as with other propositions which we know. After a while it is natural that we sometimes mistake the mere appearance of these feelings for knowledge. Now this I take to be a fairly unobjectionable empirical claim. However, I would also claim that

these feelings of certainty are neither necessary nor sufficient conditions of knowledge. That such a feeling of certainty is not a sufficient condition of 'knowing that' is amply illustrated by the numerous cases where we feel quite certain that something is the case only to find out that it isn't at all. That this feeling of certainty is not necessary is shown by the fact that we <u>can</u> quite correctly say of someone that he knows that such and such is the case, say, just after he has learned with understanding that it is so, and before any feelings of certainty set in. Of course, if the proposition is analytic or the consequence of a deductive proof, the onset of the feelings might be very rapid, but if it is an empirical proposition, they could be much slower.

4. <u>An Analysis of Rote Learning</u>

It will be recalled that I singled out four learning locutions for special study, 'learning that', 'learning why', 'learning how', and 'learning to'. I have already argued that 'learning why' is usually to be understood as a summary locution for more specific 'learning that' locutions. I then analyzed learning with understanding in terms of the 'learning that' locution. I now want to take a look at the two remaining locutions to see if either of them might serve as a paradigm for rote learning.

Now since 'learning with understanding that' implies

'knowing that', one might take a hint from Ryle and see if rote learning might be associated with 'knowing how' or 'learning how'.[7] This would seem to be a promising line since 'learning how' usually is used with reference to a skill such as learning how to type or learning how to swim, and the learning of skills is often associated with rote learning. However, one must be careful in this regard since some skills are very complex, such as learning how to do logic or play chess. The learning which goes into one of these complex skills would seem to require a great deal of understanding. However, we might take another clue from Ryle. He argues that 'knowing how' cannot be reduced to 'knowing that' and that 'learning how' cannot be reduced to 'learning that'. Intelligent practice is not the step-child of intelligent theorizing.

Now Ryle's main point must be, it seems, beyond dispute. What he is concerned to show is that the intelligent execution of a skill (knowing how) need not be preceded by a shadowy operation of dictating to oneself what to do. And in the complete context of his polemic against Cartesian dualism, this does seem to be correct. However, Ryle's claim that 'knowing how' cannot be "reduced" to 'knowing that' is ambiguous. If it means that one cannot completely specify a skill in terms of factual statements,

---

[7]Ryle, op. cit., 1949.

without remainder, then he is probably right, at least for many skills. If, on the other hand, it means that 'knowing how' entails no 'knowing that' statements, then it must be false. This is especially obvious with complex skills such as chess. It would be extremely odd to say that someone knows how to play chess if he does not know that the king can move only one square at a time in any direction except in the move called "castling." Again I am not sure that there are any basic facts which must be known in order for someone to know how to play chess, but it seems that there must be some facts. (An analysis similar to the one already given concerning learning a subject matter could be given here to show that there must be some 'knowing that' statements entailed by a correct application of 'knowing how' even if these could vary from person to person.) There is, however, a sense of 'knowing how' which does not seem to entail any 'knowing that' statements, e.g., "That one-year old child knows how to walk." This sense of 'knowing how' will be isolated and discussed in Chapter III.

Let me try to apply these ideas to 'learning how'. The situation seems to be entirely similar with respect to 'learning how' as it is with 'knowing how'. There is no "direct" entailment of any specific 'learnings that', and thus no reduction in Ryle's sense, but that there must be some entailment seems to be beyond doubt.

Let me consider in a bit more detail what it is like to learn how to swim, for example. In learning how to swim, it seems that one of the things which must be learned is to hold one's breath when one's face is submerged. Disastrous choking and coughing ensue if one does not learn this. It might be thought that what has been learned here is <u>that</u> one must hold one's breath while under water, but as Ryle points out the 'learning that' locution seems to suggest a sort of mental recitation such as "hold your breath" occurs before plunging in. Now that such a recitation must occur is patently false, and although such a 'learning that' <u>may</u> occur, it need not. But <u>something</u> has been learned. What is it? I think that the 'learning to' locution which was naturally used in describing the situation points the way out. A person need not learn <u>that</u> he must hold his breath while under water, for I have already argued 'learning that' implies 'believing that', and no such belief need be held or even suggested. But he <u>does</u> need to learn <u>to</u> hold his breath while under water. And this is, I believe, the paradigm locution for rote learning. Although 'learning how' seems sometimes to imply 'learning that' and thus beliefs, 'learning to' says nothing about beliefs nor understanding, and I think that we do not want any analysis of rote learning to imply belief of any sort. If John has learned to hold his breath

while under water, he simply does it. We might say he does it by habit, or even, without stretching usage too much, by rote. (Although the phrase, 'by rote', is most commonly used to refer to a "recitation" of some sort, it does seem to be applicable here too.)

It can now be seen that typical psychologists' descriptions of learning can be used to give an analysis of the 'learning to' locution.

> X has learned by rote to q if and only if after a series of trials, the probability that X will q on the next trial is nearly one.

A few words need to be said concerning this analysis. A "trial" is simply a situation in which the response, q, is appropriate. Thus in my swimming example, the first trial might be the first time someone puts his head under water. If 'q' is 'hold his breath when under water', then clearly this is an appropriate response in this trial, and someone has learned by rote to hold his breath when under water if and only if after a series of dunkings (the series seems to require at least one term), the probability that he will hold his breath when under water on the next trial is nearly one.

Although this analysis seems to have application to locutions like "John has learned to do arithmetic" or "Sue has learned to play chess," the intuitive meaning of a trial loses its clarity here. Thus one cannot specify

very exactly what a "trial" would be like in the following: "John has learned by rote to do arithmetic if and only if after a series of trials the probability that John will do arithmetic on the next trial is nearly one." This might be thought to be an objection to my analysis, but the objection is removed once it is noted that these locutions are really elliptical for a 'learning how' locution. Thus "John has learned to do arithmetic" is elliptical for "John has learned <u>how</u> to do arithmetic." I do not claim here that my analysis applies to 'learning how' although it will be shown later that there is a close relationship between 'learning to' and 'learning how'.

Using this analysis, I can now give an account of 'learning that' locutions in which the learning is meant to be rote learning. That is, I can analyze statements such as "John has learned by rote that the universe originated with a huge explosion" (cf. Section 3). It seems that three cases can be distinguished. First, the case in which John believes what he has learned. Second, the case in which John doesn't believe what he has learned, but does not believe its contradictory either, having simply reserved judgment. And third, the case in which John not only does not believe what her has learned, but actually believes the contradictory of what he has learned.

The first case is the easiest. John has learned by rote that the universe originated with a huge explosion,

and since 'learning that' entails 'believing that', John also believes that the universe originated with a huge explosion. Furthermore, he has learned by rote to say that the universe originated with a huge explosion. This last, according to my analysis is equivalent to saying that John has been in a series of trials and the probability that he will say that the universe originated with a huge explosion on the next trial is nearly one. The description of the trials very probably will include his having read a particular statement of the origin of the universe, his having been told it several times, and perhaps his trying to memorize one of the particular forms of words which express the proposition that the universe originated with a huge explosion. This sort of situation seems to be, unfortunately, a very common one. A student reads something or hears a proposition expressed, believes it, but doesn't have any reasons for his belief, and therefore, simply learns it by rote. That is, he learns by rote to say or write a particular form of words which expresses his belief.

The second case was the one in which John does not believe that the universe originated with a huge explosion, and yet we describe him as having learned by rote that the universe originated with a huge explosion. Since I have already argued that 'learning that' entails 'believing that', it would seem that this is a logically impossible

situation. John would have to both believe the proposition, because this is entailed by his having 'learned that', and not believe it by hypothesis. This might seem to be a counterexample to the idea that 'learning that' entails 'believing that'. However, if we examine the situation a bit more closely, I think that it will be seen not to be a counterexample.

Let me ask just what sort of a situation would lead us to say that John does not believe that the universe originated with a huge explosion and yet also say that he has learned by rote that it did. It seems clear enough that we would need some such evidence as the following: John tells us in private that he simply doesn't know what to believe about the origin of the universe. He has learned what Gamov says, but he is unconvinced. Nevertheless, he realizes that if he does not give Gamov's answer on the test, his teacher, an ardent Gamov fan, will fail him. Therefore he memorizes the answer and gives it on the examination.

The way to describe the situation now seems obvious. "John has learned by rote that the universe originated with a huge explosion" is, in this case, <u>elliptical</u> for "John has learned that Gamov (or John's teacher) <u>says</u> (or believes) that the universe originated with a huge explosion, and, furthermore, John has learned by rote to <u>say</u> that the

universe originated with a huge explosion." The belief implied by the 'learning that' in the first conjuct, Gamov says that the universe originated with a huge explosion, is no longer inconsistent with John's suspension of judgment concerning the actual origin of the universe, and the second conjunct expresses just what it is that John has actually learned by rote. Thus the inconsistency vanishes.

The third case was the one in which not only did John not believe that the universe originated with a huge explosion, he actually believed the contradictory. Nevertheless, we want to describe him as having learned by rote that the universe originated with a huge explosion. This situation is much like the foregoing. Here, too, John probably has a teacher who will insist on his giving the "big bang" answer on a test despite John's belief that this is false. The analysis might proceed along precisely the same lines as before. "John has learned by rote that the universe originated with a huge explosion" is simply elliptical for "John has learned that Gamov and John's teacher believe that the universe originated with a huge explosion." Again there is no logical contradiction now, and it is not even necessary to say that John holds inconsistent beliefs, even though this latter is empirically possible.[8] Unfortunately, it is all too common.

---

[8]The point about its being logically possible to hold inconsistent beliefs is easy to see if we schematize

It should be noted here that the dogmatic teacher is not the only way to explain the last two cases. John may simply have been required to know what the "big bang" theory was. In this case the analysis for the elliptical original expression would become "John has learned that certain people believe that the universe originated with a huge explosion and John has learned by rote to say that certain people believe that the universe originated with a huge explosion."

The important point in the cases where the person does not believe what he has been said to have learned is that if the learning is expressed by a 'learning that' locution, we have independent evidence for the learning, e.g. performance on a test, and the apparently incompatible belief, e.g. John's statement of what he actually believes. Lacking this independent evidence, we are perfectly justified in assuming that what has been learned is also believed. The independent evidence refutes this assumption by showing the apparently incompatible beliefs not to be so at all when the learning locutions are rephrased. Thus the "form" of words used in describing learning is to a

---

the situation. Thus '(X believes p) & (X believes $\sim$p)' is not a logical contradiction. However, as I have pointed out in the second example, what is logically impossible is that someone both believes and does not believe the same proposition. '(X believes p) & $\sim$(X believes p)' is a contradiction. The distinction between these two cases is easy to see in the schematic form. The difference is in the scope of the negation.

large extent immaterial, but the concepts which these words express in their paradigm uses are extremely important. I believe these last examples show quite clearly how the same locution can be used to express a variety of conceptually different situations.

CHAPTER II

FURTHER COMMENTS ON THE ANALYSES OF "LEARNING WITH UNDERSTANDING" AND "ROTE LEARNING"

1. Philosophical and Psychological Certainty

In Chapter I, Section 3, I claimed that a feeling of certainty was neither a necessary nor a sufficient condition of an analysis of 'X knows that p'. That "feeling sure' is not a sufficient condition has never been seriously denied. Only too often is our confidence misplaced. We feel sure, only to find out that we were mistaken, that we did not know at all. However, "feeling sure" has been claimed to be a necessary condition of an analysis of 'knowing that', and although I believe that my argument against this view is correct, I must expand it and try to show in a bit more detail why anyone would want to claim that "feeling sure" is a necessary condition of 'knowing that'.

In his otherwise convincing article, Norman Malcolm says,

> Whether we should say that you knew, depends in part on how confident you were. In case (2), if you had said "It rained only three days ago and usually water flows in the gorge for at least that long after a rain; but, of course, I don't feel absolutely sure that there will be water," then we

> should not have said that you knew there would be water. If you lack confidence that p is true, then others do not say that you know that p is true, even though they know that p is true. Being confident is a necessary condition for knowing.[1]

Is this correct? I am afraid that as it stands, Malcolm's claim is almost impossible to evaluate, as it is hopelessly ambiguous. In two places, Malcolm uses a form of the locution, "being confident," and in another place, the locution, "feeling sure." Apparently he takes these two locutions to be synonymous, but as I shall try to show, they can have at least three distinct meanings and Malcolm simply does not specify which of these meanings he understands his locutions to have.

To begin with, I would like to distinguish between what I shall call "philosophical certainty" and "psychological certainty." Philosophical certainty is that certainty which seems to have exercised philosophers since Plato's time and is the notion under question in the various attempts to give an analysis of the good reasons clause in the classical analysis of 'knowing that'. What is essentially involved here is the question of when the reasons are "good enough" to serve as justification for knowledge. Very roughly it seems that the reasons must be sufficiently good to be "philosophically certain."

---

[1] Norman Malcolm, "Knowledge and Belief," Mind, Vol. 61 (1952), 178-189, especially p. 179.

One form of philosophical certainty is almost certainly what is usually called "logical certainty." If the reasons for a belief are the steps in a valid deductive argument leading to that belief, and furthermore, we can simply assume the premises of the argument to be true, then it seems clear that the reasons are logically certain and then it is philosophically certain that the reasons are "good enough."

However, this sort of philosophical certainty does not take us very far. There are very few null-premiss deductions leading to interesting results. Most of our knowledge seems rather to be of the empirical variety which, although it may be cast in the form of a valid deductive argument, requires that we have some way of determining whether the empirical premises are true. And this determination must end in something other than a null-premiss deduction or tautology. Thus, the determination of what constitutes philosophical certainty for empirical judgments is the most important and least well understood problem of the analysis of knowledge, and I cannot go into this problem in any detail because it is almost as vast as all of contemporary epistemology.

What I would like to do, however, is to indicate in a very brief form some of the major lines which have been taken to attack this problem. One of the most

popular lines until recently centered around the concept of "sense data." It was argued, in one form or another, that if we press the justification of empirical reasons far enough, we eventually get to empirical evidence which is logically self-justifying. This "evidence" is supposed to be associated with the immediate contents of sense-experience--that is, one is logically sure of the having of sense data--although it is admitted that mistakes could, of course, be made in interpreting the sense data or making judgments concerning it. Thus "I see an apple" is fallible, but "red-here-now," the underlying evidence for the apple statement, is supposed to be infallible. I do not want to go into the myriad objections to this view now. Suffice it to say that in this extreme form it has been rejected by many philosophers.

Another historical approach to the problem involved the notion of synthetic a priori truths. Very roughly, this line held that there are truths which no empirical evidence can refute or confirm, but which nevertheless apply to empirical statements as the framework in which empirical evidence must be placed in order to be comprehensible as evidence at all. A classic example of such a purported synthetic a priori truth is the statement, "Every event has a cause," which, although clearly applicable to empirical evidence, cannot be refuted because every supposed

counterexample is taken not as refuting the statement but as labelling our ignorance of the real cause. Thus a seemingly causeless event simply makes us search harder for the cause of which we are ignorant. This sort of approach imparts philosophical certainty to empirical evidence by a sort of vague coherence of the structure of empirical evidence as a whole. I realize that this discussion has been quite unfair to the numerous ways in which this notion has been worked out by different philosophers, but again I can only plead the vastness of the subject.

Another variation which seems to try to combine the best features of the two preceding approaches runs something like this. We do not have any logically infallible sense data, but, nevertheless, certain empirical statements, although they could be mistaken, do in fact possess the requisite philosophical certainty. Statements such as "I am now sitting at a table writing," are of this kind. They are <u>so</u> well confirmed that questioning them would be foolish, despite the fact that future experience might lead us to reopen the question of their truth. Furthermore, these kinds of statements, since they are true on the whole, impart the necessary philosophical certainty to empirical reasons.

Now it should be noted that all of these approaches to an empirical "philosophical certainty" seem to depend

on logical certainty somewhere. Thus sense data are logically incapable of being fallible. Synthetic a priori truths are established by the logical requirement that we have a way to frame our evidence. Even in the case of basic statements which could be wrong, it seems that it is a logical point that not very many of them can be wrong.

However, I want to mention one more approach to philosophical certainty which does not seem to involve this reduction to logical certainty. Professor D. M. Armstrong of the University of Melbourne has suggested to me that this philosophical certainty need be no more than the empirical certainty associated with empirical laws. Consider the following case: Suppose a person were found who could take a look at the horses in a race and then predict unfailingly which one would win. He does not have to study their past records; he just watches the parade to the post and then picks the winner every time. Suppose further that in no case is he able to give us any further information on his method of choosing winners than that he believes a particular horse will win a particular race, and in each case he is correct. It seems to me that despite a lingering feeling of a massive coincidence, we might be compelled in such a case to admit that the man really does know that such and such a horse will win such and such a race.

But what are the salient features of this knowledge? It seems that the man must <u>believe</u> that such and such a horse will win. Let us suppose that on occasion he says that he does not <u>believe</u> a particular horse will win, but picks one anyway, and is usually wrong. That is, only when he believes the proposition is he correct, and in such cases he is always correct. The second point is that he must be in a certain situation; he must see the parade to the post. Suppose he sees the horses in other situations, but is usually wrong in his predictions in those situations. He must be in the particular kind of situation of seeing the horses parade to the post. Third, given these two conditions, it seems to be an empirical law that his belief is always true. Are not these sufficient for <u>non</u>-inferential knowledge? Armstrong claims that they are.

The major problem seems to lie in seeing how the man's repeated successes justify their being an empirical "law" connecting his belief and the conditions he is in with the truth of the belief. Laws seem to depend somehow on a theory and there is no obvious theory here. However, Armstrong also claims that ordinary knowledge based on perception is of this pattern also. If I believe that I see a table and I am in some normal state and there is a law connecting the normal state and the having of this belief with the truth of what is believed, then Armstrong

claims that I know, non-inferentially, that I see a table. Furthermore, I need not be able to specify the law--it is sufficient that one exists. In this case the "normal" situation needs to be spelled out eventually so that we can take account of our occasional errors of perceptual judgment, but the fact remains that we are right in such cases so often that such a spelling out certainly seems a plausible suggestion. Furthermore, in this case Armstrong argues that we do have at least an idea of a theory in which such a law might fit--a physiological theory connecting perception and brain states. Thus even though we do not know the form of the law, it certainly seems plausible to suggest that there is one in cases of ordinary perception. I shall have more to say on this below.

Thus Armstrong's analysis of non-inferential knowledge takes the form

X knows (non-inferentially) that p if and only if

(1) X believes that p

(2) X is in a certain condition, C

(3) Whenever anyone is in condition, C, and believes a sentence, p, of class P, then 'p' is true.

If this is right and if non-inferential knowledge is the way to stop the regress in the good reasons clause of the classical analysis of knowledge, then the philosophical certainty here is simply the empirical certainty of an empirical law. If, as a matter of fact, we are always

right about certain beliefs in certain conditions, then this would seem to be sufficient for knowledge. Such a situation would provide us ultimately with the required philosophical certainty.

Two points are immediately obvious from this brief discussion of philosophical certainty. First of all, philosophical certainty is predicated of reasons and evidence and statements expressing reasons and evidence. A statement is philosophically certain if it is "good enough" for knowledge, and this in turn means that it is supported by certain kinds of evidence which are either logically or empirically certain.

The second point is that philosophical certainty is a dichotomous concept and does not admit of degree except perhaps indirectly in empirical certainty. It is clear that this is so in the case of logical certainty. Either something is logically certain or it is not, and there are no degrees here. The case is somewhat different with empirical certainty. Here it seems that one can admit of something's being more or less probable, more or less well confirmed by the evidence. But even in this case there is a point on the continuum beyond which it would be foolish to seek further confirmation. It is at this point that empirical evidence becomes "philosophically certain," i.e. able to support a knowledge claim. At any rate, I do not

have to give a detailed analysis of empirical certainty for it to be obvious that beyond this point, wherever it is, or however it is established, the empirical evidence is "good enough." Thus the main point for my purposes is that philosophical certainty is an all-or-nothing concept. (If anyone objects to my use of 'certainty' in talking about the objective characteristics which reasons must have in order to be good enough for knowledge, I am perfectly willing to talk about philosophical "acceptability" instead.)

Now, could philosophical certainty be the kind which Malcolm is talking about in the quotation? I think it clearly is not. Even though it is not clear whether or not his "being confident" is a dichotomous concept, it is abundantly clear that it is predicated of <u>people</u> and not of reasons or evidence. Furthermore, if confidence is predicated of people, it must be obvious that it admits of degrees in a quite straightforward way. One can be more or less confident, feel more or less sure of one's evidence. It is thus quite clear that philosophical certainty and psychological certainty are prima facie different concepts. The former is dichotomous and predicated of evidence. It is in a sense "objective." The latter admits of degrees and is predicated of people. It is a "subjective" concept.

Now these two notions may be related somehow, but

they have no direct relation. An example or two should serve to make this clear. One can _feel_ very confident of something for the weakest of reasons or even with no evidence at all. On the other hand, some people, especially people at the frontiers of knowledge, may have excellent objective reasons for a result and yet perhaps because of the unusualness of the result, not feel very confident at all. I can and do admit that _some_ notion of philosophical certainty is involved in the analysis of knowledge. However, I do not think that this is what Malcolm intends.

The question can now be asked if Malcolm's claim is true if he means that _psychological_ certainty is a necessary condition of knowledge? Unfortunately, this is _still_ an ambiguous claim; for it seems that one can distinguish two different kinds of certainty _within_ psychological certainty. These are an "emotional certainty" and a "volitional certainty." The former is usually expressed by such locutions as "I feel sure of it," or, "He feels quite confident of that." Thus emotional certainty is connected with how we _feel_ about certain things. On the other hand, volitional certainty is connected with what we are willing to _do_ or _say_. This certainty is commonly expressed _without_ the terminology of feelings, e.g. "I _am_ very confident that the sun will rise tomorrow." I want to argue that emotional certainty is neither a necessary

nor a sufficient condition of knowledge and that volitional certainty, properly understood, is an element in simply believing something to be the case. Thus I shall argue Malcolm's claim is false if it taken to refer to emotional certainty and true, but misleading, if it is taken to refer to volitional certainty.

2. Emotional Certainty and Claims to Knowledge

Let me first consider that kind of psychological certainty which I have called emotional certainty. It will be recalled that this is a predicate of people, admits of degrees, and is expressed in the language of feelings and emotions.

The way to show that emotional certainty is not a necessary condition for knowing is to distinguish, as Woozley and White have done, between knowing and claiming to know.[2] A person can justifiably claim to know that something is the case without its being the case that he really knows it at all. The justifiability of the claim is distinct from the truth of what is claimed. Just as someone can justifiably believe something which is not the case, so can one justifiably claim to know when one in fact does not know. A few examples should make this clear.

---

[2] A. D. Woozley, "Knowing and Not Knowing," Proceedings of the Aristotelian Society, Vol. 53 (1953), 151-172; and A. R. White, "On Claiming to Know," Philosophical Review, Vol. 66 (1957), 180-192.

A student of Ptolemaic astronomy probably believed that the sun revolves around the earth. Furthermore, he may have had good reasons for his belief. He had a theoretical system which indicated this and furthermore it probably looked to him as if the sun revolves around the earth. For all that he knew, the sun could have revolved around the earth. Now surely we might want to say that he was correct in _saying_ or _claiming_ that he knew that the sun revolves around the earth, although we would want to deny that he actually _did_ know that the sun revolves around the earth; for it doesn't.

Or to take another example, a blind man relying upon his friend's usually accurate testimony could properly _claim_ to know that snow is black if his friend told him so, although, of course, he doesn't know it at all.

What these examples show, I think, is that being justified in making a claim to knowledge does not entail that the claim is true. A person can be justified in _claiming_ to know that such and such is the case without knowing that it is. Now my position is that emotional certainty may be a necessary condition of a proper _claim_ to knowledge but not of knowledge itself. However, I have not yet shown this. All my examples have shown so far is that claims to knowledge and knowledge are to be distinguished. I think my examples have also shown that a

justifiable claim to knowledge does not entail knowledge, but the question remains, does knowledge entail being justified in claiming to know? For if it does, and if emotional certainty is a necessary condition of justifiable claims to knowledge, then it seems that emotional certainty would after all be a necessary condition of 'knowing that'.

That emotional certainty is a necessary condition of a justifiable claim to knowledge seems clear enough. We often say such things as, "How can you possibly _say_ you know that? You don't even feel sure of it," meaning by this to question the propriety of his _claiming_ to know it.

To show 'knowing that' does not _entail_ a justifiable claim to knowledge is much harder, however, for _as a matter of fact_, whenever someone knows something, he is almost always also in a position to claim to know it. Thus examples are hard to come by. One way of attacking the problem is by showing that in so far as claims to knowledge are simply _claims_, and the truth of what is claimed does not entail the justifiability of the claim, knowledge does not entail the justifiability of a claim to knowledge and thus not emotional certainty. This argument is, I think, correct; for we sometimes make wild claims which just happen to be true, but for which we have no justification whatever. But this is not enough.

For while the argument could be admitted, it could be objected that given a complete analysis of what it is to be justified to claim knowledge, the reason the entailment from the knowledge to the justifiable claim to knowledge fails is because knowledge does not imply some of the conditions <u>other</u> than the emotional certainty one. It still <u>does</u> entail emotional certainty. I think this countermove is wrong because all that I can conceive of as an analysis for justifiably claiming to know something would be the following:

X is justified in claiming that he knows that p if and only if

(a) X believes that p

(b) X has a high degree of emotional certainty that p

(c) If the reasons for X's belief were true, they would be good reasons for believing that p.

But it can easily be seen that the only condition which could fail to be entailed by knowledge that p is the emotional certainty one, (b). But perhaps the analysis is incomplete.

Another tack suggested by White's paper is to show that the reason "I know that p, but do I feel sure?" is queer is <u>not</u> because "I know that p" entails "I am sure that p." The reason that this particular conjunction of statements is odd is because as soon as the first conjunct is uttered, there is at least a prima facie case for saying

that I am justified in my claim to knowledge. As soon as I say "I know," if I am being sincere, I believe that I am justified in the claim I am making. But the second part of the conjunction then asks if I really am justified after all. If I really doubted the justification of my claim, I shouldn't have made it, and ordinarily I do not. The conjunction is odd, not because of any logical contradiction in what is claimed, but rather in the fact that I am put in the odd position of claiming and wondering about the justification of the claim at the same time.

The case is much clearer in the third person where this sort of oddness does not arise. Consider the following: James has been writing his dissertation under the direction of Professor Smith and is now in the process of taking his oral examination. Suppose Professor Jones asks Jim if the sentential logic is complete. Jim is very nervous and in general manifests the signs of feeling very uncertain about the answer. It seems to make perfectly good sense for Professor Smith to say, "Jim knows that the sentential logic is complete, he just feels very unsure of himself." Let me take the case even further. Suppose after a great deal of hesitation, Jim said that the sentential logic was indeed complete. This in itself might not convince Professor Jones who might believe Jim was simply guessing. But suppose that Jim then proceeded to prove

the completeness of the sentential logic still with extreme nervousness and great signs of insecurity. Surely, despite Jim's nervousness and feeling unsure, Professor Jones would have to admit that Jim did know it. Jones might later say to Smith, "He knows his stuff all right, but he must get over feeling so insecure if he is to make a favorable impression. Maybe the pressures of an oral examination tend to interfere with the ability of candidates to say what they really know." What Professor Jones would be indicating in such a remark is that feelings of insecurity and uncertainty are often due to sources other than not knowing the material. I think that this shows quite conclusively that feeling sure is not a necessary condition of knowing, although as a matter of fact, when one knows, one also usually feels sure.

It is precisely this sort of third person case which I used in the first chapter to deny the necessity of emotional certainty in an analysis of knowing. In that case it was a student beginning a course who had learned with understanding that such and such was the case, and thus knew it, but who had yet to attain those feelings of certainty which continued study of the course material and use of the knowledge as a matter of fact almost invariably bring.

One last point. It seems to make perfectly good

sense to say, "How can you possibly say that you know that? You aren't even sure of it," meaning by this not to question whether or not he knows it, that could be admitted, but only the propriety of his claiming to know. But if knowledge entailed justification for claiming to know, this would be logically impossible. Furthermore, if Armstrong's suggestions concerning non-inferential knowledge are correct, then examples could quite easily be constructed in which the person knows that such and such is the case, but is not justified in claiming to know it; for all that is required for him to know it is that he believe it and there be a law connecting his belief and state with the truth of what he knows. He need not feel certain at all.

3. Volitional Certainty and Belief

I want now to turn to the second species of psychological certainty, that species which I have called "volitional certainty." It will be recalled that I claimed that this kind of certainty is properly expressed without the language of feelings and is connected with what we are willing to do or say. The difference between emotional certainty and volitional certainty can be brought out by saying that my use in the first chapter of 'believes that' as a generic term to cover 'hopes that', 'fears that', 'suspects that,' etc. simply will not do. A person can hope that something is the case although he believes its

contradictory. A Republican in 1964 might have hoped that Johnson would be defeated, but in his heart of hearts he believed that Johnson would be elected. I want to suggest that it is emotional certainty which is present in the 'hopes that' case and absent in the 'believes that' case while volitional certainty is present in the 'believes that' case and absent in the 'hopes that' case. Then since 'believing that' is a necessary condition of 'knowing that', volitional certainty must also be present in knowledge.

H. H. Price gives the following analysis of belief.

> So far we have distinguished the following four factors in the situation called "believing p":--
> (1) Entertaining p together with one or more alternative propositions, q and r.
> (2) Knowing a fact (or set of facts) F, which is relevant to p, q, and r.
> (3) Knowing that F makes p more likely than q or r, i.e. having more evidence for p than for q or r.
> (4) Assenting to p; which in turn includes
> (a) the preferring of p to q and r;
> (b) the feeling of a certain degree of confidence with regard to p.[3]

I want to make several observations on this analysis. First, Price is analyzing a kind of belief which he admits others might call opinion or reasonable or justifiable belief. This is made clear by conditions (2) and (3) which I would largely take account of in my analysis of knowing in the good reasons clause. Clearly (3) is

---

[3]H. H. Price, "Some Considerations About Belief," Proceedings of the Aristotelian Society, Vol. 35 (1934-1935), 229-252, especially pp. 236-237.

directly concerned with giving an account of good reasons, i.e. of why it is more rational to hold p than some alternative. (2) is not so directly concerned with the "goodness" of reasons for a belief, but it nevertheless seems to deny the possibility of entirely unsupported beliefs. I only want to make two points concerning this condition. First, even if beliefs must have some evidence relevant to them, the requirement that we know the evidence seems much too strong. Surely belief in some evidence would be strong enough for mere belief. Of course, the stronger requirement may be necessary for rational belief, but that is not the kind of belief I am discussing.

Secondly, I am not at all sure that mere belief requires any evidence relevant to the belief, even if the evidence is only believed. The only argument Price gives for this condition is the example of a fork in the road where we have no evidence as to which fork to take. Nevertheless, we do choose one fork, and in so doing, Price argues that we only "act as if" we believe what we have chosen. I readily admit that such cases exist and that the "act as if" locution is the appropriate one to describe them. However, I deny that this is the case in all situations in which we have no evidence. The reason the "act as if" description is appropriate in Price's example is because the person has not actually assented

to the proposition. This will become clearer in the following.

At any rate, it seems perfectly plausible to say that someone believes in God, or life on other planets, even though he does not believe there is any evidence relative to the truth or falsity of the belief. The difference between believing and merely acting as if one believes lies in the chance nature of the latter as opposed to the far-reaching constant influence of the former. We take the right-hand fork and merely act as if it is the correct way because given another similar situation we might equally have taken the left-hand fork. But surely we are not just flipping a coin when it comes to believing in God or in life on other planets. Even if we do not believe that we have any evidence relevant to the belief, there may be other reasons, not evidential, but perhaps emotional which push us to one side or the other. The empirical difference between these two notions lies, I think, precisely in this chance nature of "acting as if." A person acts if a proposition were true if, given similar situations, there seems to be a merely chance variation in what he does. If the variation is not chance, then we can say that he believes the choice he makes.

Now condition (1) is simply meant to say that one can understand a proposition without judging it true or

false. In order for me to believe a proposition I must understand the language in which it was expressed to me, know the meanings of the words involved, and so on. This is all that Price or I mean by "entertaining" a proposition. Price uses the form "thinking of A as B" without believing or disbelieving that A is B, without knowing that A is B or not. This is, I think, perfectly non-controversial. Furthermore, in most cases the alternative proposition to p is simply ~p. Of course, "entertaining a proposition" or "grasping the sense of a statement" undoubtedly needs further analysis.[4] However, the main point seems clear enough. We can grasp the meaning of a statement independently of judging the statement to be true or false. Part of this "grasping" may indeed involve understanding the truth conditions of the statement, i.e. the ways in which we would go about deciding its truth or falsity, but this does not mean we actually must go through such a procedure.

Let me know turn to condition (4), assenting to the proposition. It is obvious that (4.b) simply expresses the condition of what I have called emotional certainty, i.e. the feeling of a certain degree of confidence. Of course, I have to deny (4.b), or else either my analysis of knowledge in terms of belief is wrong, or my arguments against

---

[4]As an example of such a further analysis see P. F. Strawson, "On Referring," Mind, Vol. 59 (1950), 320-344, reprinted in Essays in Conceptual Analysis, edited by A. G. N. Flew (London: MacMillan and Co., 1956).

emotional certainty as a necessary condition of 'knowing that' are wrong. In fact, however, entirely similar considerations of claims and what is claimed hold good mutatis mutandis for belief as well as for knowledge. Thus emotional certainty may be a necessary condition of justifiable claims to belief, but they are not necessary conditions for the belief itself. Actually it is not even queer in the first person to say "I believe that such and such is the case, but I'm not at all sure of it." "But mustn't there be some feeling of confidence," it might be objected. "Of course, I don't have to feel sure, but certainly I must have some feeling of confidence in order to believe something." I admit that we almost always and perhaps always do, as a matter of fact, have these feelings. But that does not make them a part of the meaning of belief, any more than the fact (if it is a fact) that all creatures who possess a heart also possess a kidney makes the possession of a kidney part of the meaning of 'possessing a heart'. The only semblance of an argument which Price gives for (4.b) seems to be the following:

> The other case is that in which there is no evidence at all on either side, as when we come to a fork in a road, and have no evidence as to which of the two branch-roads is the right one. Here it seems to me that assent, and therefore belief, is impossible. We can only decide to act as if road A was the right one. But we do not on that account believe it. For we feel no confidence about its being the right one.[5]

---

[5]Price, op. cit., p. 236.

Now I have already claimed that what this argument shows, if it shows anything, is that (2) is a necessary condition for belief. But the argument, even if it is valid, and I have argued it is in general not, does not rest, as Price makes it rest, on the lack of a feeling of confidence. Rather it rests on the lack of a lasting commitment to the belief, a commitment made on a basis other than mere chance. The man merely "acts as if" not because he does not have any feelings of confidence, but because he does not prefer one fork to the other. And he does not prefer one fork to the other because he has no reason (evidence or emotion) for so doing. He merely "acts as if" and feelings of confidence may or may not be present.

Again because as a matter of fact feelings of certainty almost invariably accompany belief, it is very difficult to construct a counterexample to Price's condition (4.b). However, I believe that the following is a counterexample. Suppose that an ingenious philosopher were to take the sceptical position concerning the external world and proceed to destroy a student's confidence in the existence of the external world. The student might very well say, "All right, I don't feel at all sure that there is an external world. You have made me feel very uneasy about this, but nevertheless, I believe that there is an external world." Would we say with Price that the student

has merely decided to act as if there were an external world? I think not. It would be extremely odd to deny that the student actually does <u>believe</u> that there is an external world. There is no element of chance here.

But condition (4.a), the preferring of p to its alternatives (to ~p as I would say) <u>is</u> a necessary condition of belief, and it is the one which enables me to explain how Malcolm's claim can be true but misleading. It is true because belief is a necessary condition of knowledge. It is misleading because it appears that volitional certainty is applied only to knowledge and not to belief as well. Price characterizes preference as an all or none sort of thing. We either prefer p to its alternatives or we do not. We decide that p. We come down on one side of the fence or the other and there is no question of degree here.

Now although preference itself may be an all or none affair, it does not seem to me that volitional certainty is. Surely there are degrees of volitional confidence. I am <u>quite</u> confident that the sun will rise tomorrow and all my actions show this extreme confidence. For example, I plan on doing things tomorrow, I do not alter my present activities, etc. On the other hand, I am not nearly so confident that the San Francisco Giants will win the National League pennant this year although I believe that

they will. However, it seems to me that one can use the notion of betting and accepting bets at varying odds as expounded in modern subjective probability theory to give an explication of what it is to have sufficient volitional certainty to be said to believe a proposition. I suggest that all the essential sense of "preferring p to ~p" can be captured in the following explication:

> X prefers p to ~p if and only if X would be willing to accept even money odds on p, but would not be willing to accept even money odds on ~p.

This explication I think rules out mere "acting as if" and also shows the point in the scale of confidence at which the all or none concept of preference comes into play.

Unfortunately, the explication will not quite do as it stands. I would have to add a condition to the effect that X was acting rationally in accepting or refusing the bets, i.e. was basing his decision solely on what he believed the evidence for p to be. My original example of the Republican who hoped Johnson would be defeated but believed he would not shows the necessity of a rationality condition. If offered an even money bet on Johnson's being defeated, my man would probably accept it; for to not do so would show that he didn't really believe the Republican candidate, Goldwater, would win, and even though he did not believe Goldwater would win, it is simply not good form to bet against the man one supports. Thus

because of other than evidential considerations, he would irrationally accept a bet he did not believe he would win. But even so, the size of the bet might make the man act rationally in spite of himself. Although he would undoubtedly bet on Goldwater if the stakes were comparatively small, if he were somehow forced to accept an even money bet on one of the two candidates at stakes which were very meaningful to him, I suggest he would probably bet in accordance with his belief, and not in accordance with non-evidential considerations. However, these non-evidential considerations do not affect the logical analysis of the degree of confidence a person has in a proposition in order to believe it.

I think that another way of putting this point is to say that belief is essentially dispositional. Part of what believing a proposition means is that we are disposed to act in certain ways because of our belief. Now to call a term "dispositional" is not by itself to admit that there are any directly stateable laws connecting the having of the disposition with what would happen in certain circumstances. If John believes that Kennedy was a good president, we would expect him to say so in certain circumstances. We would expect him to support many of Kennedy's policies and decisions and many other things as well. But this does not mean that if we know that John believes that

Kennedy was a good president, we could unerringly predict that John would do these things in the way that we can predict that my pencil would fall if I released it. Rather, if John failed in a very great number of cases to do those things which we expect of someone who believes that Kennedy was a good president, then we would have reason to doubt that he believed it, although there could almost always be special circumstances which would explain the failure of our expectations and yet be consistent with the claim that John believes that Kennedy was a good president. One such explanation could be that John doesn't understand English and all the occasions in which we expected him to behave in certain ways and were disappointed were such that he would have had to have understood English to behave in the expected manner. Another might be that on all the occasions of frustrated expectations John believed that the person being referred to was not Kennedy, but someone else, because he did not realize that 'JFK' and 'Kennedy' referred to the same person.

Ryle seems to say that dispositions are very much like laws.[6] They enable us to explain, predict, and many other things that laws also permit us to do. In fact according to Ryle there are only two differences between dispositional statements and laws. The first is that

---

[6]Ryle, op. cit., pp. 117-125.

while laws are "open at both ends," dispositional statements are "open at only one end" for they mention a particular person or thing. Thus while laws are true of all the objects of a certain class, dispositional statements are ordinarily true only of particular objects. This is quite obvious in the case of beliefs which we ascribe to particular people.

The second difference is that although both laws and dispositional statements are expressible in terms of subjunctive conditionals, dispositional statements are not so easily disconfirmed as laws. For example, the law, "All unsupported objects fall when released in a gravitational field," is meant not only as a report of past occurrences, but also as a prediction of possible future occurrences, both actualized and unactualized. Thus we get the subjunctive, "If any unsupported object <u>were</u> released in a gravitational field it <u>would</u> fall." Furthermore, this is usually taken to be refuted by any counter-instance, unless the counter-instance can be explained away in terms of other laws. But consider "John believes that Kennedy was a good president." The subjunctive form of this is much harder to state since such a belief involves John in being prepared to do a great number of different things. Furthermore, this set of things is quite vaguely defined, which in itself leads to much less rigor

insofar as a counter-instance being taken as a refutation. But the situation is even less rigorous than this, for Ryle seems to imply that even if the set of things which John must be prepared to do were precisely specified, a certain minimum amount of failure to perform would not refute the ascription of the belief. Dispositions are essentially less precise than laws.

I think that Ryle is correct in this, but unfortunately the very close connection between dispositional statements and laws can easily lead one to believe that dispositional statements are really just first approximations to laws. To begin with, one is tempted to collect a number of instances of dispositional statements which differ only in the particular individual named and generalize away this reference to particulars. Thus from observing that A, B, . . .all believe such and such and then do thus and so, we conclude that anyone who believes such and such would do thus and so. Secondly, it is all too easy to assume that the greater vagueness in a dispositional statement need not be essentially so. In order to make a dispositional statement into a law, we merely need to spell out all the relevant circumstances in more detail, and we will eventually get a law describing just the same sorts of entities and events which the dispositional statement originally described.

I think that this sort of a procedure would be a mistake, albeit a natural one. It may be that a proper analysis of any dispositional statement entails the existence of a law, but this is not to say that the form or even the concepts used in the law are at all similar to the form of the dispositional statement or the concepts used in it. Nelson Goodman has shown the extreme difficulty of giving any sort of straightforward analysis of subjunctive conditionals, and subjunctive conditionals are most often used as the means to give a first approximation to an analysis of dispositional statements.[7] Thus if I were to say "This match is ignitible," the corresponding subjunctive might be "If this match were struck, it would light." But this latter defies analysis in any straightforward way as Goodman has so adequately shown. Nevertheless, we are convinced that there must be some law, even if we can't state it, which supports this subjunctive. And, of course, there very probably is. Only the form of the law would undoubtedly be much more complicated than "Whenever X, Y follows" and, of even more importance, the concepts used in the law would probably not refer to things like matches' being struck and lighting at all. Rather it might use concepts such as oxidation, specific quantities of heat, temperatures, friction producing heat, etc.

---

[7]Nelson Goodman, Fact, Fiction, and Forecast (Cambridge, Mass.: Harvard University Press, 1955).

The case is entirely similar with belief. To ascribe a belief to John does indeed involve ascribing a disposition to him to behave in a certain range of ways, and this in turn is to say that a subjunctive of the form "If John is in a certain condition, then having the belief that he does, he would probably do such and such an act," is true. Furthermore, we can even say that there is some law which supports this subjunctive and which could be used in some kind of explanation of the behavior, but this explanation need not at all be the usual kind of explanation in terms of reasons, beliefs, and desires. We need not have a clue as to what the law might look like.

Roughly, the reason straightforward generalizations of the subjunctive conditionals connected with dispositional statements won't do as analyses of the dispositional statement is that they are not laws since they are not ultimately confirmed by their instances. And the reason these subjunctives are not confirmed by their instances is that they usually use concepts to describe the events which are not appropriate for being used in laws. However, these events can sometimes also be described in terms which are appropriate for being used in laws and thus the dispositional statements are supported by laws. We have no laws connecting matches, being struck, and lighting, but we may very well have laws connecting oxidation, specific

heats, and heats of friction. Similarly we have no laws (or very few) connecting beliefs, intentions, and actions, but we may very well have laws connecting brain states, brain processes, and muscular or neurological movements. I shall return to this point in Chapter IV where I shall propose criteria for evaluating proposed analyses of dispositional statements--those analyses which are so popular in reductionistic psychology. (By "reductionist psychology" I simply mean those schools of psychology which attempt in one way or another to eliminate talk about reasons, beliefs, intentions, i.e. mental concepts, in favor of talk about observable things.)

However, what I have said so far shows explicitly how certain common sense ideas concerning testing for learning and the limitations of testing are well grounded. First it shows that since a great deal of learning involves belief, learning is dispositional in the way that belief is dispositional. And this in turn, because people who have certain beliefs tend to behave in certain verbal ways, among others, shows why verbal tests are relevant to finding out if learning has occurred. On the other hand, since there do not exist at the present very many, if any, known laws supporting the dispositional statements, testing is limited. There could always be a reason for a wrong answer on a test which would counteract the presumption that the

student had not learned that particular bit of information. If there were any known laws connecting belief and behavior, this possibility logically could not arise in those cases. The best that we can do at present is to see if the student consistently fails to behave in the ways in which a correct application of the belief statement to him would lead us to expect him to behave. If he does fail to so behave, then we have good, although not conclusive, reasons for believing that he has not learned what he was supposed to have learned.

One of the most important, although not the only set of subjunctives associated with belief as a diposition is concerned with verbal behavior. Thus if someone believes something, we suppose that if he were asked certain questions in certain circumstances, he would say or write certain things. Now the fact that belief has these "verbal" subjunctives associated with it enables me to distinguish roughly between "cognitive" learning and the learning of "skills." For it has probably already been noted that not only are learnings where belief is involved dispositional, but the learning of skills is also dispositional where beliefs are involved, if at all, in a secondary way. If I say that Mary has learned how to type, I am ascribing a disposition to Mary just as much as if I had said that Mary has learned that the Declaration of Independence was

signed in 1776. The difference between them is that a test of the former disposition tends to be of the nature of seeing if Mary can perform certain non-verbal behavior, e.g. taking a typing speed test; while a test of the latter is usually made by seeing if Mary can perform certain verbal tasks, e.g. answer questions on a test.

Of course, the distinction is not hard and fast. Mary's having learned how to type may include some dispositions to verbal behavior as well as her simple typing performance, e.g. identifying parts of the machine by name, being able to read copy, etc. But this verbal behavior surely is subordinate to her typing. On the other hand, Mary's having learned that the Declaration of Independence was signed in 1776 may involve some dispositions to perform certain non-verbal behavior, e.g. going to a dance in 1976 celebrating the 200th anniversary of the signing of the Declaration of Independence. But again this would be subordinate to her verbal behavior.

Finally one last important characteristic of belief, which will be amplified in detail in the fourth chapter, is the fact that belief is intensional or provides an opaque context. Discussions of what this amounts to abound in recent philosophical writings and Quine and Chisholm have given excellent presentations of the technical details of

the intensionality of belief.[8] For my purposes it is sufficient to summarize the intensionality of belief by saying that the most important considerations for talking about a person's behavior are not those which bear on what the situation he is facing in fact is like, but rather those considerations which bear on what he believes the situation he is facing to be like.

This intensionality of belief can be illustrated in a graphic way by considering the by now famous example of the morning star. Let me suppose that John is endowed with at least a modicum of logical acumen and thus

(a) John believes that the morning star is identical with the morning star.

but, of course,

(b) The morning star = the evening star = Venus.

Thus by substitution we get

(c) John believes that the morning star is identical with the evening star.

But (c) may not be true and in fact at certain times in history certainly was not true. But this is a paradox. By following well-accepted rules of logic, we have derived a false conclusion from true premises. Clearly the trouble is not with substitution but with this particular kind of

---

[8]Willard Van Orman Quine, Word and Object (New York: The Technology Press of The Massachusetts Institute of Technology and John Wiley and Sons, Inc., 1960); and R. M. Chisholm, Perceiving: A Philosophical Study (Ithaca, N.Y.: Cornell University Press, 1957).

context. This situation has been called the opacity of belief contexts and has given rise to a huge amount of recent philosophical discussion which it is beyond my purposes to consider. It is sufficient to notice that the ordinary substituting of names or descriptions of the same thing does not necessarily preserve truth value in all contexts.

Let me consider another example more suited to my purposes. Suppose a student has learned that the sum of the squares of the legs of a right triangle is equal to the square of the hypotenuse and furthermore has learned how to prove it. Then he is faced with the question on a test to state and prove the Pythagorean theorem. But suppose that for some reason he doesn't believe that the Pythagorean theorem is the same as the one he has learned. It seems that the opacity of belief gives some weight to the student's oft-voiced complaint, "Oh, I knewhow to do that, but I didn't think that was what was wanted (or meant)."

And another example. Suppose a student wants to learn "how to think properly," and thus takes a course in modern logic because he believes that logic will teach him how to think. However, if someone substituted for 'logic' in his belief a description of the study of the propositional calculus, the functional calculus, formal proofs, formal systems, etc., it seems highly unlikely that the

student would also believe that would teach him how to think properly.

Thousands of similar examples could be given. It is important to note this opacity of belief contexts and thus by means of the analyses I have given the opacity of many learning contexts; for many problems in learning take on a new light when it is realized that 'learning' often appears in opaque contexts. I shall also show in the next section that the opacity of belief is crucial to an understanding of the good reasons clause in my analysis of knowledge and, thus, also to learning with understanding.

There are two main reasons for spending so much time on getting clear on just what sort of certainty is connected with knowledge and in what ways. The first of these reasons has already been suggested in my discussion of the student taking an oral examination. What the examiner wants to know is how good a student's evidence is for his beliefs. That is, the examiner ought to be concerned only with philosophical certainty--the relation of evidence to beliefs--and whether the student has such evidence. Unfortunately, testing situations are often quite anxiety producing, and they sometimes cause great feelings of insecurity in the student. The student is emotionally uncertain of himself, and this sometimes interferes with his performance, despite the fact that he is usually

volitionally certain of what he knows, and, furthermore, his reasons are often philosophically certain--adequate for his belief. At the very worst the examiner sometimes is tempted to conclude from the obvious emotional insecurity of the student that he has not learned with understanding what he is supposed to have learned.

On the other extreme there is the case of the brashly confident student who exhibits a great deal of emotional certainty, but whose beliefs are unsupported by evidence which is philosophically certain. In this case the danger lies in the other direction. The examiner is tempted to conclude that the student knows what he is talking about when he may merely have cultivated an air of confidence. Of course, these points are psychologically clear, but a proper attention to the logical relations of the different kinds of certainty to what is known, believed, learned, justifiably claimed, has, to my knowledge, never been sufficiently emphasized. If learning and the testing for learning is to proceed in a valid way, the correct understanding of the kinds of certainty and what they show is indispensable.

The second major point concerns the attempts of various philosophers in the pragmatic tradition to define truth as roughly that which works. These philosophers, and they abound in schools of education, have been duly

impressed with mankind's fallibility. Only too often do we feel sure of something, only too often are we confident of something, only to find out that it is not so at all. Impressed by human fallibility and by the occasionally striking differences in the conceptual schemes of certain other cultures, these philosophers have despaired of ever knowing what the "real" truth is. As a result they have denied all "absolutes" wholesale, including truth, and have sought to define knowledge in a way that seems to be nothing more than what I would call being justified in claiming to know.

Of course, one must not be hasty in condemning someone's belief simply because it does not agree with one's own ideas, and, of course, numerous intellectual and moral outrages have been perpetrated by groups claiming to have some special insight into absolutes, and of course, one must try to understand opposite points of view. But none of this means that there is no clear meaning of 'is true', and that we all ought not to strive to believe just those things which are true. Such redefining of relatively clear concepts such as 'true', 'knowledge', 'certainty', can lead only to confusion. What I hope to have shown with my analyses is that there are perfectly good senses of these words, and one ought to be clear in using them. And if one does this, then all of the problems with which

pragmatic philosophers concern themselves can be formulated in a straightforward way without the dangers of misunderstandings and misinterpretation. Let philosophers be concerned with varying points of view, or "frames of reference" as they are sometimes called, but one can talk about this in terms of justifiable claims to knowledge without talking about 'true from a point of view'. Indeed, let us discuss beliefs which are strongly held and acted upon by certain cultures. Let us also discuss the reasons for these beliefs and try to see if they are true, but there is no reason to call these beliefs "knowledge" if they are not true. Only confusion can result from thus tampering with perfectly good notions. Let us indeed be suspicious of any claims to special insight into knowledge, but let us not reject the publicly observable and testable knowledge that we do have. I hope that my analyses have shown how one can talk about the important problem of learning without resorting to the specialized terminology of so many pragmatic philosophers of education.

4. The Having of Good Reasons

In this section I will not be able to give an account of what it is to have "good" reasons for a belief, since that is beyond my competence, and even if it weren't, it would fill several volumes. The problem here is that, in a sense, giving an account of the good reasons clause

in my analysis is what much of contemporary epistemology is all about. For instance, I believe that almost all the recent books on perception, e.g. Chisholm, Armstrong, Ayer, are attempts to explicate what it is like to have good reasons for a knowledge claim based on perception.[9] In addition they often try to clear up philosophical problems relating to our use of perceptual locutions in themselves, but the attempt to relate perception to the foundations of empirical knowledge as well is clearly evident.

Another example of an attempt to explicate good reasons is the recent development of inductive logic, exemplified by Carnap.[10] This theory attempts to give an account of rational belief, the confirmation of hypotheses by evidence, and several other important notions. And although this theory is at present only able to deal satisfactorily with very limited languages, it is already so mathematically complex as to require at least a book for its proper exposition. There has also been some thought recently about extending confirmation theory to cover value judgments; for the problem of how evidence confirms hypotheses seems remarkably similar, at least formally, to the problem of weighting reasons for and against certain

---

[9]Chisholm, op. cit., 1957; Ayer, op. cit., 1956; and D.M. Armstrong, Perception and the Physical World (London: Routledge and Kegan Paul, 1961).

[10]See for example, R. Carnap, The Logical Foundations of Probability (Chicago: University of Chicago Press, 2d. ed., 1962).

proposed actions which could be judged right, wrong, blameworthy, obligatory, etc.

However, we need not await the eventual success of these programs to use correctly locutions like 'learning', 'knowing', etc.; just as men reasoned correctly long before Aristotle. What I should like to try to do, then, is give some sort of analysis of simply having a reason for a belief, without considering at all how particular reasons for or against a belief come to be combined into good or conclusive or justifying reasons for or against the belief. The analysis I will give of someone's having a reason for believing a proposition will thus be compatible with his having other reasons for not believing the proposition or even for believing the contradictory. In fact it will be compatible with his knowing that the contradictory is the case, i.e. with his having good reasons for believing the contradictory. Thus a statement need not even be true for someone to have a reason for believing it, and this, I think, corresponds with our intuitions concerning the having of reasons. For we often say, "I admit that you have reasons for believing that, but nevertheless it just isn't so."

Despite these apparently narrowing restrictions on my proposed analysis, it should be evident as I proceed that the task of analyzing the mere having of a reason is not trivial, and further if the analysis is successful, it

will shed valuable, although partial, light on the previous analyses of learning and knowledge.

In his provocative article Quine gives a hint as to what might be involved in an analysis of having a reason for a belief. He says,

> The totality of our so-called knowledge or beliefs from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. Reevaluation of some statements entails reevaluation of others, because of their logical interconnections--the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having reevaluated one statement we must reevaluate some others, which may be statements of logical connections themselves. But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole.[11]

Now Quine is talking about what must be done to accommodate a bit of evidence into our conceptual scheme, when for some reason or other the evidence or experience is unexpected or contrary to some of our previously held beliefs. And he argues forcefully, I believe, for the

---

[11]W.V.O. Quine, "Two Dogmas of Empiricism," From a Logical Point of View (Cambridge, Mass.: Harvard University Press, 1953), pp. 42-43.

view that there is no logical reason why we must change one part of the conceptual scheme rather than another. The considerations are more pragmatic. What parts would be simplest to change? What changes would cause the least disturbance? Which might be most fruitful for future predictions and explanations?

Admittedly the notion of our conceptual scheme as a net or field impinging on experience at the periphery is a metaphor, but it is an extremely suggestive one. It suggests a way of saying at least metaphorically what it would be like to have a reason for a belief. One has a reason for a belief if one can fit it into the conceptual scheme in the appropriate place and see just how it is connected with the remainder of the scheme. This means in a sense that a person's entire conceptual scheme and its interrelations constitutes the reasons for any belief of his.

Nevertheless, for practical purposes, certain relatively small subfields of the total conceptual scheme can be singled out as particularly relevant to any given belief and the showing of the connections of the belief with the rest of this subfield does serve us in practice as a sort of "giving the reasons for the belief." After all, as Aristotle pointed out, we cannot give reasons for reasons indefinitely. We must start (or end) with something which

is accepted for the purposes of the particular discussion, although what is accepted can, of course, vary from discussion to discussion. For this reason I shall analyze the locution 'X has a reason, R, for believing that p' where the reason is to be taken as a certain subset of the beliefs of the person's total conceptual scheme deemed particularly relevant to the belief in the context of a given discussion. It need not be the only such subset.

From this, two conditions of the analysis are immediately apparent. First, the person must <u>believe</u> all the members of the subset; for we would hardly allow that someone had a reason for a belief if he did not even believe the reason. Second, it seems that the person must believe that the reason is evidence for the truth of the belief. I think that this captures the metaphorical notion of a person's being able to place the belief in its proper place in the subset of reasons and to exhibit its relationships with the other members of the subset of his total conceptual scheme.

It is extremely difficult, however, to give any *further analysis of believing that a reason is evidence* for the truth of a proposition and I shall not attempt to do so. What I shall do is to try to make clearer some of the things which are and are not connected with this notion. First of all, it seems that for someone to believe that a

reason is evidence for the truth of a proposition, he must have in mind some manner in which the evidence is connected to the proposition. He must have some notion about the various ways in which certain statements are evidence for others. I do not want to include these beliefs as to the ways in which evidence is connected to statements in the reason itself. These beliefs are rather the general background giving sense to what it means to be evidence in general and not in any particular case. Some examples ought to help make this clear.

Suppose Paul believes that he is mortal. Suppose further that he believes all men are mortal and he is a man. And finally suppose that he believes that the latter beliefs are evidence for the truth of his belief that he is mortal. Thus Paul has a reason, for believing that he is mortal. Now it is also fairly obvious that the reason Paul believes his reasons are evidence for his mortality is that he believes valid true deductive arguments are evidence for their conclusions. This is the way in which he believes his reasons to be evidence for his mortality. It is also clear why this last belief in the evidential value of true valid deductions is not to be taken as a part of his reasons for the original belief. In most situations the way is taken for granted and is thus not evidence for believing the proposition but rather is evidence

for the belief that the reasons are evidence for the truth of the original belief. Of course, if someone asked Paul why he believed his reasons were evidence for his mortality then he might very well respond with his belief concerning the evidential value of deduction, but that would be moving the discussion to a different level. But that he must have a way in mind seems clear. We would hardly attribute to anyone even a belief that certain statements were evidence for the truth of another statement if he had no idea how the former were evidence for the latter. Rather we would say that he didn't even know the meaning of 'evidence'.

The fact that the way in which statements are taken as evidence is usually simply assumed illustrates, I believe, how we can and do limit the range of beliefs which are taken as evidence for another belief. The fact that we can, by moving the discussion to a different level, inquire as to the way in which someone takes a set of statements to be evidence for the truth of another statement shows how ultimately the entire conceptual scheme can be used piece by piece to justify a belief.

Another example. Suppose Paul believes that the Yankees will win the 1965 American League pennant. Suppose his reason for this is that he believes that the Yankees are as good a ball club as they were in 1964 when they won the pennant. In this case the way in which Paul believes

his reason to be evidence for his belief is that the past performance of baseball teams is good inductive evidence for their future performance. In this case Paul might also have a reason for believing that the Yankees won't win the pennant. For example, he might believe that the Baltimore Orioles who almost beat the Yankees last year are greatly improved. Paul would have to evaluate the weights of these two reasons, and perhaps others, in reaching a final judgment.

Finally suppose that Paul believes that a vixen is a female fox. Suppose he also believes this because of his further belief that the definition of 'vixen' is 'female fox'. Again Paul has a reason for his belief, and fairly clearly the way in which he takes his reason to be evidence is that statements giving the definitions of words are true in virtue of the meanings of the words involved.

Two points arise from these examples. First, the "ways" in which statements are evidence for another statement include the familiar deductive, inductive, and definitional frameworks. These frameworks give the structure necessary for calling anything evidence at all. These are the ways most commonly taken for granted even if the person who takes them for granted could not give a very precise statement of them or perhaps no statement at all. It would suffice that he consistently argued in accordance

with these structures. Second, calling the framework of evidence into account, as I have mentioned in my examples, moves the discussion to a new level. One is asked why he thinks that a particular statement is evidence for another and the answer is in terms of one of the three ways illustrated. This way of changing the level of discussion is different from a much more common way in which the discussion is changed. In the common case, the evidential framework is assumed, i.e. the questioner admits the relevance of the reason propounded, but questions its truth. In such a situation one then gives other reasons for the original one continuing until the questioner is satisfied or the level is changed to the evidential framework itself or the defender simply finds it pointless to argue any further; for the questioner seems not to accept the most obvious reasons. This last case can often be a disguised form of questioning the evidential framework itself, or it can be indicative of someone who simply doesn't understand the language.

So far I have claimed that (1) X believes every member of R, and (2) X believes that R is evidence for the truth of p, are necessary conditions for an analysis of 'X has a reason, R, for believing that p'. There are, however, still some problems. First, the way in which a person believes that a reason is evidence for a statement

may be obvious and well-accepted as it was in my examples. But it need not be. Suppose instead of the usual deductive, inductive, or definitional evidential frameworks, a person had the following framework. He believes that a set of statements is evidence for another statement when and only when a little voice tells him so. It would seem that my two conditions so far would not rule our such a person's having a reason for believing something. But intuitively we would want to say he had (in some strong sense of 'had') no reason at all.

A second problem might arise if the set of reasons were too small. For example, suppose Paul believes that Jones owns a Ford and also believes that Jones' father bought a Ford last year. Now if Paul, believing that the latter is evidence for the former, offers the latter as a reason for the former, we would probably say that we didn't see how 'Jones' father bought a Ford last year' could possibly be evidence for 'Jones owns a Ford', and we would ask Paul why he thought it was. And this question is really two questions in one. We would be asking Paul either to give us more information so that it would be apparent how the new amplified reason was evidence for his belief, or else explain to us the manner in which he takes the original reason to be evidence. And, of course, the first question of the disjunction is the one which is usually

answered. Thus Paul might say, "Oh, Jones' father gave that Ford he bought a year ago to Jones yesterday." The new reason would be that Jones' father bought a Ford last year and gave it to Jones yesterday.

These examples show the need for another condition on my analysis of 'X has a reason, R, for believing that p'. The condition I propose is 'R is evidence for the truth of p in the way X believes R to be evidence for p'. Now it is a semantical point, but one which I want to make clear here, that I take a set of statements being evidence as sufficient to guarantee the truth of each and every member of the set. Thus my newly added condition entails the truth of the members of the set of reasons, R. If anyone objects to this usage of 'evidence', I do not care, but then he will simply have to add to my analysis the condition that the members of the set of reasons be true.

That such a requirement of the truth of the reasons is necessary is easy to see. It might be thought that the mere fact that a statement follows from a valid deduction would be some evidence for its truth. But if that were the case, we would have some evidence for the most ridiculous statements imaginable; for any statement can be made the conclusion of a valid deduction if we have no restrictions on the premises. For example,

All rabbits are equations
The Easter Bunny is a rabbit

∴ The Easter Bunny is an equation

is a valid deduction, and if my analysis of someone's having a reason for a belief does not entail that the reasons are true, then there would be some evidence for the conclusion that the Easter Bunny is an equation, which is absurd. Similar absurd examples could be constructed for other types of evidential relations.

Thus my final analysis is as follows:

X has a reason, R, for believing that p if and only if

(1) X believes every member of R
(2) X believes that R is evidence for the truth of p
(3) R is evidence for the truth of p in the way that X believes that R is evidence for p.

Following the pattern established earlier of distinguishing the analysis of a locution from the analysis of the justifiability of asserting the locution, I also propose the following analysis:

X is justified in claiming that he has a reason, R, for believing that p if and only if

(1') X believes every member of R
(2') X believes that R is evidence for the truth of p
(3') If every member of R were true, R would be evidence for the truth of p in the way that X believes that R is evidence for p.

In this last analysis, (3') replaces (3) of the previous one. The use of the subjunctive preserves the intuitive feeling that even to be justified in claiming to have a

reason, the person's purported evidence must have the form at least of actual evidence.

It will be noted that this last analysis probably lets in the odd reasons that certain primitive peoples have, according to anthropologists, for some of their beliefs. I think in most cases we would intuitively want to say that these people were justified in claiming to have reasons, although we would want to deny that they did actually have them in many cases. I do not think that this analysis would allow that someone was justified in claiming to have reasons in the Easter Bunny case mentioned above. The reason for this is that subjunctives range from the obviously true to the obviously absurd, and one of the ways in which we evaluate their truth is tied to just how easily we can imagine the antecedent conditions. Thus in the case of "if that match were struck, it would light," what we are asked to imagine is quite obvious and fairly well circumscribed and the subjunctive is clearly true.

On the other hand, in the Easter Bunny example, we are asked to imagine rabbits being equations, and we are not at all sure what this might be like. Are we to imagine equations as flesh and blood or rabbits as abstract entities? What might a contrary bit of evidence for the claim that all rabbits are equations be like? It seems that (3') would not be true in the Easter Bunny case for we have <u>no</u>

idea what we are asked to imagine. I admit that there are hard cases between the extremes of the match lighting and the Easter Bunny being an equation--cases in which we are simply not sure whether the subjunctive is true or not. This is perhaps why the foregoing is not proposed as an analysis of the truth of subjunctive conditionals, but rather as an indication of how we do go about determining their truth. But there are also cases of people claiming to have fairly wild reasons for beliefs in which we are also not sure whether they are justified in their claim or not; and the virtue of the subjunctive account of these cases is that the variations are precisely the same. The closer the subjunctive is to the case of the match lighting, the more apt we are to say that the person is justified in claiming to have a reason. The closer the subjunctive is to the Easter Bunny case, the less apt we are to say that he is even justified in claiming to have a reason for his belief.

Both of the foregoing analyses rule out the problem cases noted earlier. They rule out the "little voice" example because although the reasons may be evidence for the belief, they are not so in the odd way that the person believes them to be. They rule out the "too small reasons" example because 'Jones' father bought a Ford last year' is not and would not be by itself evidence for the truth of

'Jones owns a Ford'.

It should also be noted that my analysis is neutral as to the question of whether or not the set of reasons must contain laws, lawlike generalizations, or statistical generalizations. Whether it must or not is a question to be answered in the analysis of 'R is evidence for the truth of p', and this is a problem with which confirmation theorists are dealing. It is also at the heart of the controversy concerning the deductive model of explanation.[12] My analysis has the virtues of being able to accomodate any of the conflicting views on this matter and of showing just how the problem of what constitutes evidence for a belief is related to the having of a reason for a belief. It also shows several other interesting features of the latter which are more to my purposes.

In the first place, 'learning with understanding that' and 'knowing that' have different levels or at least different degrees. The reasons which I have for a belief about modern physics are of a far different order from those of a theoretical physicist and also from those of a high school physics student. Yet it could be that we have all three learned with understanding, and, hence know

---

[12]See, for example, C.G. Hempel, "The Function of General Laws in History," reprinted in Readings in Philosophical Analysis, edited by H. Feigl and W. Sellers (New York: Appleton-Century-Crofts, 1949), pp. 459-471; and William Dray, Laws and Explanation in History (London: Oxford University Press, 1960).

that $e = mc^2$. The simple fact that we have all learned this with understanding does not entail that we all have the same kind of understanding. There are still vast differences between the understanding of the master of a subject and that of the apprentice. And conversely it shows that a person can learn with understanding without being a master of the subject, just so long as the reasons are good ones and are his. This seems to be where the "grasping a proposition" and the "opacity of belief" come in as well. The student of a subject can hear his teacher say that q is a reason for believing r, but if the student doesn't know the meanings of the terms in q, then he cannot believe q since he cannot even entertain q (see Section 3 of this chapter), and thus he cannot have a reason, q, for believing r. However, it might be that q can be recast into the form q' where the terms refer to the same things that they did in q but are now understandable to the student. Now we could say that the student has a reason, q', for believing that r.

The analyses also show, I believe, the point of psychological research to determine more precisely what the conceptual apparatus at various stages of development in the child is like. If we wish a student to learn with understanding, it does no good to attempt to get him to parrot the reasons of a modern physicist for a belief about

physics if he (the student) doesn't understand those reasons--if they are not a part of his conceptual scheme.

And I believe that the analyses show clearly a kernel of truth in the progressive education idea of concentrating on the students' problems and not on a packaged course made up by the teacher. The packaged course too often gives its reasons in the teacher's terms and not in the students', and if this is correct, my analysis shows that there can be no learning with understanding since the necessary conditions of having reasons which the student can fit into his conceptual scheme are lacking. The problem method on the other hand has at least the advantage of couching the discussions in terms which the pupil can understand--can fit into his conceptual apparatus--although it certainly does not guarantee success.

One final implication is that the analyses show that the successful teacher is the one who can not only set forth his points in a clear and lucid way, but who can give reasons for them in terms which the student can understand. Thus given the lack of perfect teachers who can support their lectures in precisely the terms all students can understand, the analysis shows the superiority of tutorial type teaching where the student by searching and probing, and the teacher by questioning and recasting points, can eventually be reasonably sure that there has been learning

with understanding. This is, of course, only the old argument favoring discussion type teaching, but given new depth by the analyses. There is nothing intrinsically better about seminar teaching than lectures. It is just that in practice the former seems to have a better chance of satisfying the conditions of learning with understanding than the latter. The virtue of the analyses is to point out what any teaching method must take into account if the pupil is to learn with understanding from it.

5. The Sufficiency of Knowedge for Learning with Understanding

It will be recalled that in the first chapter all I claimed in giving an analysis of learning with understanding was that knowledge was a necessary condition of the learning. Is it also sufficient? I think not, and I shall be concerned in this section with showing knowledge is not sufficient for learning with understanding. I shall also examine attempts to add a condition which would jointly with knowledge be necessary and sufficient.

The classical problem with the analysis that I have given of 'knowing that' is that in explicating the good reasons clause, it always seems to be the case that for the reasons to be as "good" as necessary, they must themselves be known, and this in turn leads to an infinite regress since we must explain what it is to "know" a reason. But

this leads to an infinite regress only in case the type of knowing involved in knowing the reasons is always of the form which itself requires reasons. Can we not stop the regress by simply showing that there is a form of knowledge which does not require the giving of reasons--some ultimate or immediate or non-inferential form of knowledge? If this is the case, and if we can give an independent account of this non-inferential knowledge, then it seems the original account of what can now be called inferential knowledge can be saved after all. In any case it is clear that whatever the character of this non-inferential knowledge, it is not of the kind which can be learned with understanding. Rather it is the basis upon which all learning with understanding rests. It makes possible learning with understanding.

I must guard against a possible misinterpretation of what I have just said. All I have said is that if someone knows something non-inferentially, then he cannot learn it with understanding, and very probably cannot learn it at all. It is something that he just knows. I have not said that he cannot learn with understanding how his non-inferential knowledge is connected to the rest of his conceptual scheme nor even what the logical structure of non-inferential knowledge is like. Indeed, I would want to insist that he can learn these latter things with understanding. Perhaps an example will make this clearer.

Suppose that my knowledge that I am now sitting at a desk writing is non-inferential. This does not entail that my knowledge that this is non-inferential is itself non-inferential. The sentence 'I know that I am now sitting at a desk writing' may be a report of non-inferential knowledge, but that does not entail that the sentence "I know that the sentence 'I know that I am now sitting at a desk writing' is non-inferential" is itself a report of non-inferential knowledge. The latter sentence almost certainly is inferential, simply because philosophers are always trying to give reasons for or against it. The distinction is closely connected to that of object language and meta-language.

It should now be clear why my account of learning with understanding in terms of inferential knowledge is relatively independent of the controversies surrounding the proper analysis of non-inferential knowledge, e.g. the controversy between phenomenalists and realists. First of all, the reasons which are commonly given and accepted in response to the question, "How do you know that?", are themselves ones for which justification could be asked. It is only if they are continually challenged by a philosopher that one finally reaches non-inferential knowledge. And this is commonly not done in deciding when someone has learned something with understanding. In fact this is

just the same point that was made in the previous section where I urged that the way in which someone believes a set of statements to be evidence for another statement was not in itself to be taken as a member of the set, but rather the way gives the background information commonly accepted. This "manner" can be questioned, but that moves the discussion to a different level.

Secondly, although there must be an acceptable account of non-inferential knowledge for my analysis to be correct, it is immaterial what precisely that account is, for it is of non-inferential knowledge, whereas my account of learning with understanding is in terms of inferential knowledge, and one can, in practice, almost always settle all the questions concerning learning with understanding before reaching the point of considering non-inferential knowledge.

Thus, I might be tempted to claim that the whole class of non-inferential knowledge is never learned with understanding nor even learned at all. Nevertheless, I cannot give sufficient conditions for learning with understanding by simply claiming that all inferential knowledge is learned with understanding. The objection is that what one knows non-inferentially may be what another can know only inferentially. Thus if Paul knows something non-inferentially, he cannot have learned it with understanding,

but Sam may know it inferentially and thus could have learned it with understanding. However, this need not surprise us, and in fact I think that it is quite in accord with common usage.

For example, it seems quite plausible that what a sighted person knows non-inferentially might be known only inferentially by a blind man. The sighted person knows in general non-inferentially what a person looks like, but the blind man, if he has not been without sight from birth, knows this only inferentially from touch and the relating of his tactile sensations to his remembered images of what people look like. The surprising thing here is that it follows from my analysis that the blind man may have learned with understanding that a person's face has such and such a look while the sighted person has not. Upon reflection, however, this need occasion no alarm. The blind man needs to concentrate his whole attention on gaining a knowledge of what his friend looks like while the sighted person takes it for granted. For example, the case is analogous to the esthetic experience of having lived in the mountains all of one's life and then suddenly noticing how beautiful they are. One knew that they were beautiful all along, but only when one has stopped and for some reason really related their appearance to one's conceptual scheme does one learn with understanding how truly beautiful they are.

Thus since non-inferential knowledge cannot be specified independently of the particular knower of it, I cannot give a general account of just when a proposition was learned with understanding and when not. But again this never bothers us in practice, for we can almost always tell for any particular person and particular proposition whether or not his knowing the proposition is a case of learning with understanding or not.

But there is yet a further problem. Professor D. M. Armstrong has suggested to me that even in the case of inferential knowledge, or at least in the special class of inferential knowledge which he calls knowledge <u>based</u> on good reasons, it may not be correct to analyze the "having" of reasons in the way that I have done. A weaker sense of "having" may be sufficient. It might be sufficient merely that the belief in the analysis of knowledge be acquired as a causal result of the good reasons. The reasons themselves may not have to be believed. He gives as an example early experiments on psychical phenomena where the subject <u>knew</u> certain things about what the experimenter was thinking, but knew them as a causal result of intakes of breath, facial expressions, bodily movements, etc. on the part of the experimenter without consciously noticing these clues and probably without believing that they were clues.

Another example might be that of a scientist who makes an intuitive guess that such and such is the case.

Suppose that he is unable to give good reasons for this belief. We might say only that he believes that such and such is the case, but it also seems that still being unable to give good enough reasons, he might come to know that such and such is the case inferentially, i.e. a story could be constructed to convince us that he not only believed, but knew as a causal result of certain of his other scientific knowledge or experiments, despite his inability to give any reasons. Now I would want to claim that the scientist has not learned with understanding. In fact he would quite probably say, "I know that such and such is the case, but I don't understand it." Then suppose that the scientist worked out the necessary theory and experiments not only to prove his belief but fit it into its proper place in scientific theory. Now he does understand it and we could say that he has learned with understanding that such and such is the case. It is his newly acquired ability to give his reasons which enables us to say that he has learned it with understanding.

These examples would seem to deny the necessity of the conditions that a person believes the sentences in his set of reasons and believes that they are evidence for the truth of his belief in my analysis of a person's having a reason for believing something. I think that these examples are counterexamples to any analysis of knowledge

such as I have given *if* that analysis is taken to be an analysis of *all* forms of 'knowing that' or even of all forms of inferential knowledge. It seems that it is sufficient that there be some sort of causal connection, which the person need not necessarily be aware of, between the reasons and the belief that is known. However, it is also the case that these merely causal types of knowledge are *not* the ones which are entailed by 'learning with understanding that'. In order for someone to have learned something with understanding he must not only "have" the reasons in this unconscious sort of way, but "have" them in the sense that he is aware of them in the way in which I have sketched above and illustrated with the case of the scientist when he has found the theoretical support for his hypothesis. Neither the scientist with the intuitive hypothesis only, nor the psychic can be said to have learned anything with understanding. So it seems that the analysis of knowledge which I have given might more appropriately be termed an analysis of "rationalizable" knowledge. At any rate the analysis of knowledge I have given in terms of what may be called the strong sense of "having" reasons, does seem to be appropriate to learning with understanding.

Let me now consider in some detail the following example of rationalizable knowledge which yet apparently fails to have been learned with understanding. It seems that

it could be the case that Brown knows that he is married in the sense of rationalizable knowledge. He believes it truly and has good reasons for this belief in the strong sense of 'have' which involves (although does not entail) his being able to give these reasons. I think it would be odd to say that he has learned with understanding that he is married, although he must, of course, have learned other things with understanding, e.g. what marriage is, what rituals must be gone through, etc. But consider the classic comedy situation in which Brown wakes up in the morning after a wild party to find a strange woman in bed with him who claims to be his wife. Upon investigation he learns that he did indeed marry this woman the night before and even gradually begins to remember the events of that wild night. Surely in this case we would not find it odd to say that Brown has learned with understanding that he is married. But what is the difference between the two cases? Why should we be willing to apply learning to the one but not the other? The only reasons I can see is that in the second case Brown forgot that he had gone through the ritual which he remembered in the first case. But if one can learn something that one has forgotten, there seems to be no logical bar to his having also learned it before he forgot it. The reason it is odd to say that Brown learned with understanding that he is married in the first case, but not in

the second, is rooted, I believe, in a concpetion of learning as something passive. One learns by having knowledge poured in. But the learning stops once the actual business of living begins. We then merely apply our learning to situations in which we find ourselves. Thus in the first case according to this conception, Brown has learned what would be sufficient conditions for him to be married and then recognizes that he is in such a situation and forthwith knows that he is married. Of course, in the normal case he also anticipates being in the situation and often participates in a rehearsal of the ritual. So when the actual event occurs, it is usually not the case that he finds out anything he did not already know or anticipate even though he may have some sort of unanticipated emotion of one sort or another at the actual ceremony. Thus we tend to deny that Brown has learned that he is married in the normal case because everything seems to have been learned and practiced before. There is no element of surprise, at least on one level. Nevertheless, we would not deny that he learned what it is like to actually go through the ritual instead of merely imagining or practicing it. He also learned what it is like to actually set up housekeeping with his wife and so on.

It seems then that the mere familiarity with what it takes to be married is the reason we do not in the

normal case say that Brown has learned that he is married. There is no element of surprise, little that was not anticipated. The trouble is in trying to formulate just what is absent in the normal case that is present in the waking-up-in-the-morning case. There seem to be two alternatives here. One is simply to take the hard line and claim that even in the normal case Brown has learned with understanding that he is married. There is much to recommend this view since obviously a great deal of learning has gone on over a long period of time in order for Brown to know that he is married and this learning is very probably with understanding. The problem is to see just how this prior learning makes it correct to say that Brown has learned with understanding that he is married. The other, "soft," line would be to insist that common usage and our intuitions are correct, and that what cannot be learned in the normal case can be learned in the comic case because of some added element in the latter. The most likely candidates for this added element, surprise, or coming to be aware of what you were not previously aware of, seem, however, to be inadequate; for we are not always surprised at what we learn and coming to be aware of what we were not previously aware of seems to characterize the normal case as well. In any example of knowledge we at some time become aware of what we were not previously aware of, and limiting the time of

coming to awareness seems wholly arbitrary.

I think that I am somewhat attracted to the hard line and I want to say just one or two further things in defense of it. This sort of situation in which we easily say that someone knows that such and such is the case and yet find it odd to say that he has learned it seems to be one in which a person in a sense "makes" the proposition true by an action of his. Now without getting involved in the problems of future truths, I think that this raises some interesting points. Consider the case of a boy who begins learning what the institution of marriage is like and what the sufficient conditions are for being in such a state. At some point we can probably say that he has learned with understanding that it would require such and such for him to be married. Suppose that later he comes to believe that he will eventually marry. I think we could say that he has learned with understanding that there will be a time, t, such that at any time after t, it would be true to say of him that he is married. Then suppose he and a girl set the date. The time, t, has now become definite, say t_o. The date comes and goes and he is now married. What I want to say is that the reason "he has learned with understanding that he is married," is odd is simply because it occurred a long time ago and in successive stages. The oddity comes in the fact that before t_o the present tense

is inappropriate for describing what he has learned and thus knew long ago. He decided to take the action which would set the time definitely at which the present tense would become applicable to something he had learned long ago. Even if after setting the date and girl, he backs off, it is still true perhaps that he has learned that there will be a time, t, just not t_o, at which time he will be married. Or he may learn with understanding through his unfortunate courtship that he will not get married after all. The comic case can be explained by saying that he had simply forgotten that he had set a definite time and girl and then he learned that he actually had done so. He did not learn with understanding that he was married in this case in any different way than in the normal case. Rather he simply relearned that he had set the time and girl.

I think this sort of argument might make the hard line of saying even in the normal case that Brown has learned with understanding that he is married more palatable. However, I admit that it certainly takes liberties with the ordinary grammatical forms, and I am not wholly satisfied with it. To conclude then, although I see no totally satisfactory way of adding to my analysis to give sufficient as well as necessary conditions for 'learning with understanding that', I must repeat what I have already said so often; that in particular cases, in particular situations, we are

almost never in doubt whether a piece of knowledge has been learned or not. It is rather the understanding of it with which we are concerned, and of this part, I claim to have given an informative analysis.

6. Further Considerations on Rote Learning

I want now to consider several problems with the analysis I have given of rote learning. This examination will, I believe, show how an apparently clear definition such as is often given by psychologists in cases of this kind turn out, on analysis, to be vast oversimplifications.

It will be recalled that my analysis of rote learning was

> X has learned by rote to q if and only if after a series of trials, the probability that X will q on the next trial is nearly one.

One of the first problems occurs in my having inserted the notion of a "trial" in the analysans without putting it into the analysandum. Now this procedure is philosophically sound so long as the concepts used in the analysans are at least as well understood as those on the analysandum, or serve to tie several important notions together. This procedure is unacceptable if the concepts used in the analysans are more problematic than those in the analysandum or are problematic in the same way. Now this is precisely what is wrong with "trial" in the foregoing analysis. I previously said that it was a description of a situation in

which the response, q, was appropriate. But for quite a number of situations this notion has no clear application. For example, consider instead of "hold his breath while under water" the related response "hold his breath." The analysis now becomes "X has learned by rote to hold his breath if and only if, after a series of trials, the probability that he will hold his breath on the next trial is nearly one." But what could a trial be like here? The problem is that learning to hold one's breath is a skill which can be used in many <u>different</u> trials. It would be more natural to say "X has learned <u>how</u> to hold his breath," and possessing such a skill, he can exercise it in many "appropriate" situations; when under water, when asked to, when passing a chemistry laboratory, etc. The problem is that rote learning of a skill occurs only in very highly structured situations. It has very limited transfer value to a variety of situations and perhaps that explains the attempt to <u>define</u> rote learning in transfer value terms in the first chapter. Although, as I have argued, this characteristic is not sufficient to differentiate totally between rote learning and learning with understanding, it would seem necessary to include it in the analysis of rote learning. Thus the analysis would become

> X has learned by rote to q in S if and only if X has been in a series of S's and the probability that X will q in the next S is nearly one.

In this analysis 'q' and 'S' refer to a class of responses and a class of situations and not to particular members of these classes.

But there is another problem that arises. Although I have stipulated that when I use 'has learned', I mean to exclude the possibility of forgetting, I have not done so in the analysans. This problem was taken care of automatically in the case of learning with understanding because if someone knows that such and such is the case, then by the nature of knowing, he cannot have forgotten it. But the case is much different where the analysans contain only terms of a generally behavioristic nature. It could happen that at the present X has learned by rote to q in S and yet it be the case that the probability he will q the next time he is in S is very small indeed, since the next time he is in S may be years away, by which time he will have forgotten. In fact this sort of difficulty infects almost all behavioristic analyses of any kind of human action. The problem is that the specified behavior simply need not occur. (See Chapter IV.) I see no other way of patching up my account to take care of this problem than by using a subjunctive. The analysis now becomes

X has learned by rote to q in S if and only if

(1) X has been in a series of S's
(2) The probability that X would q if X were now in S is nearly one.

But the analysis is still incomplete; for consider

the following case. Jim has had several dunkings, but continually comes up choking. Then someone explains that attempting to breathe with his face under water is what causes him to choke. He should hold his breath when under water and then he won't choke. Jim understands this explanation, and as a result of the explanation, not the dunkings, he resolves to hold his breath the next time he is dunked and suppose that the probability that he would do so if he were to be dunked now is nearly one. It seems that both conditions are satisfied and yet we would not want to say that Jim has learned by rote to hold his breath while under water. He learned it all right, but as a result of understanding what was going on. We would be more apt to say that Jim has learned with understanding that if he wants to avoid choking, he must hold his breath while under water. So I must add yet another condition to my analysis to the effect that the reason the probability is what it is is due to the trials and nothing else. So the analysis becomes

X has learned by rote to q in S if and only if

(1) X has been in a series of S's
(2) The probability that X would q if X were now in S is nearly one.
(3) The probability mentioned in (2) is a causal result of the "trace" left by (1) and nothing else.

By "trace" I mean whatever is physiologically and cognitively effected in the person by his experiences in

the trials. I need such a concept to eliminate possible wild counterexamples which might arise if the causal chain were to go outside the person and somehow yet cause him to do what he has learned although he had forgotten it in the meantime. "Trace" is a "that which" concept which will ultimately require some sort of empirical specification. Logically it serves the following purposes: The trace in rote learning is <u>that which</u> is caused in the person by his experiences in the trials and it is <u>that which</u> enables him to perform the action he has learned.

I am now in a position to show how my analysis of rote learning differs from a merely conditioned response. To begin with, the action, q, is an <u>action</u> and not mere physical movements. That is, whatever has been learned is exercised intentionally by the learner. His learning is under his voluntary control. He can choose whether or not to perform. As I shall use the term, 'conditioned response', this element of intentional control is lacking. The subject is conditioned to make the response no matter <u>what</u> he intends.

Even in the case of rote learning, except for special stipulations of the meaning of 'learning' by psychologists, I do not think that anyone would want to call a merely conditioned response not under the agent's control 'learning'. Rather we would say that he had been

made to respond in such a way. Of course, my analysis of rote learning might also do quite well as an analysis of conditioned responses if the execution of what has been inculcated were not under the person's intentional control. And I think that this accords quite well with our intuitions concerning rote learning. Rote learning is indeed a minimal form of learning and it shares important elements with conditioned responses. It occurs through practice and in specific situations. The only difference is that its execution is intentional--it is an action and not a mere movement. Moreover, the person who has learned by rote to do something almost always believes that he is in a situation in which the execution of what he has learned is appropriate whenever he is, in fact, in such a situation. Perhaps these two cognitive elements, intention and belief, were what Hilgard was referring to when he said that all learning is to some extent cognitively controlled. (See Chapter I, Section 1, footnote 2.) If this is so, then I agree wholeheartedly, but as I have tried to show, this alone does not make the distinction between rote learning and learning with understanding one of degree only.

The changes in the analysis of rote learning do not substantially alter the remarks I made in Chapter I, Section 4, concerning the rote learning of propositions. Thus "John has learned by rote that the universe originated

with a huge explosion" might become "John has learned by rote to say that the universe originated with a huge explosion whenever asked about the origin of the universe by his odd teacher." The only change is in specifying the specialized situation which led one in the first place to say that John had learned the fact by rote. And this requirement was certainly made use of in my earlier discussions of this sort of situation. The part about John's other beliefs or lack of them simply goes through unchanged.

I think these analyses show the kernel of truth contained in the incomplete attempts of Chapter I to find a difference between rote learning and learning with understanding. I have already mentioned the transfer of training approach and pointed out that rote learning is indeed tied to rather specific situations, and I have reflected this in specifying the situation in both analysans and analysandum. The mistake was in assuming that since rote learning does have this essential reference to a specific kind of situation, the difference between rote learning and learning with understanding lay solely in this fact. That there is more involved I have already argued.

The "practice" approach of Chapter I also has its kernel of truth. Rote learning must be the result of a certain amount of "practice" in what is learned and even

more important, it must be due solely to this practice. I believe that this requirement is reflected in my third condition. The mistake here was to assume that since practice was essential to rote learning, it could be made the basis for distinguishing between rote learning and learning with understanding. This won't do, however, since the amount of practice required can vary even within rote learning, and learning with understanding may also require practice of some sort. Furthermore, the important part of the practice conception--that it and it alone cause the increased probability of performing--was completely lacking in the discussion in the first chapter.

In discussing the modifications which had to be made to my original analysis, I think another important fact was illustrated. This was that rote learning and learning with understanding are most often combined in ordinary situations. It seems that in the usual case one does not learn to hold one's breath while under water either wholly by rote or wholly with understanding. The usual case combines the two. Thus someone might learn with understanding that if he tries to breathe while under water, he will choke, and this learning helps him to learn to hold his breath under water. But as Ryle pointed out, after he has learned to do this, he need not recite to himself each time before he does it, even though he could explain why he does

hold his breath. He does it by rote even though he did not wholly learn it by rote. And it would be a rare person indeed who could learn to hold his breath under water solely by listening to instructions and explanations. Part of the learning is usually due to the explanation and part to the practice; however, the logical distinction between the two kinds of learning remains, and a further analysis of their interrelations must await the next chapter.

7. A Derivative Sense of Learning with Understanding

With the aid of some of the notions and analyses of this chapter I am now in a position to defend more fully the requirement that the truth of what is learned is a necessary condition of learning with understanding. In order to see the necessity of the truth condition let me consider in some detail the objection I mentioned in Chapter I. I there supposed that someone might object to the truth condition by pointing out that we would certainly want to say of someone that he has learned with understanding about Ptolemaic astronomy despite its being false. In fact I also supposed that it might even be claimed that someone could learn with understanding that the sun revolves around the earth.

There are two lines which can be taken in answering this objection. First, I could claim that "Bill has learned with understanding that the sun revolves around the earth" is simply elliptical for "Bill has learned with understanding

that Ptolemaic astronomers believed or said that the sun revolves around the earth." This reformulation preserves the truth condition and in certain cases, especially in the case of present day students studying the history of science, this would be a most natural thing to say. The student of the history of science has as reasons for his various beliefs, historical reasons, i.e. reasons for believing certain historical figures held certain beliefs. He in general does not have reasons which would lead him (the present day student) to believe that the sun revolves around the earth. Rather his reasons are reasons for believing that historical figures claimed to have their own reasons for certain of their beliefs.

However, it is important to notice that this line is not limited to the present day student. If Bill lived before Copernicus' time, he also might have learned with understanding that Ptolemy and his followers <u>believed</u> that the sun revolves around the earth. His reasons would be of the kind which included his having been taught this, having read it in books on astronomy and so on. Of course, this line works with the ancient student of astronomy only in case he reserved judgment on whether the sun revolves around the earth or believed that the sun does not revolve around the earth. Only in such cases would his reasons be only of the sort which would lead him to believe that this

is what people said.

But what of the student who not only learned that certain people believed that the sun revolves around the earth, but also believed this himself and believed that he had good reasons for his belief. Can we not say that he has learned with understanding that the sun revolves around the earth even though this is not true? This brings me to the second line of attack. I have been at pains in this chapter to distinguish being justified in claiming to know something and actually knowing it. I have pointed out that although knowing something entails the truth of what is known, a person can be justified in claiming to know what is false. I have also argued that 'learning with understanding that' entails 'knowing that' in the sense of rationalizable knowledge. What I now want to suggest is that there is a derivative sense of learning with understanding which entails being justified in claiming to know but which does not necesssarily entail knowing, and thus does not entail the truth of what is learned. In this derivative sense of learning with understanding the person must then believe what he has learned and have good reasons for the belief although the belief need not be true.

Thus my hypothetical student of ancient astronomy could in this derivative sense have learned with understanding that the sun revolves around the earth. He

believed it and furthermore he had good reasons for it. One of these good reasons may have been that it looked to him as if the sun revolves around the earth and moreover he had a good common sense theory of the relation of the way things look to the way they are, although in this case he went wrong in interpreting his theory. Another reason might have been his mastery of Ptolemaic astronomy which not only was able to explain this one fact, but many others as well.

I have several reasons for calling this sense of learning with understanding, "derivative." First of all, I think that being justified in claiming to know is derivative to knowing in the sense that it is through understanding what knowing something is like that one understands how one could be justified in claiming to know even if one doesn't really know. If this is so, then learning with understanding which entails being justified in claiming knowledge must be derivative to that learning with understanding which entails knowledge itself.

Secondly, believing a proposition with good reasons (the derivative sense of learning with understanding) seems to gain its <u>point</u> through the <u>possibility</u> that the proposition is actually true. That the derivative sense of learning with understanding has what value it does have seems to be dependent on the consistency of what is learned

with the rest of what the student knows (and is therefore, true). The derivative sense of learning with understanding is most often used in cases where what is learned is possibly true for all that the learner actually does know at the time.[13] The reason that my ancient astronomer may have learned with understanding (derivative) that the sun revolves around the earth is because it was possible that the sun did revolve around the earth for all that anyone knew at that time.

This notion of understanding in terms of what was known at a certain time also seems to me to be the main point behind the contrast between "internal" and "external" criticisms of a theory. The internal criticism of a theory consists in putting oneself in the shoes of a believer in the theory, in imagining what he might have learned with understanding in the derivative sense.[14] On the other hand, external criticism involves critisizing a theory in terms of what has by now been learned with understanding and is, thus, true.

The third reason for including the truth condition in the primary sense of learning with understanding is quite

---

[13]See Jaakko Hintikka, Knowledge and Belief (Ithaca, N.Y.: Cornell University Press, 1957), for a related discussion of "It is possible, for all that a knows, that ."

[14]Compare this with R.G. Collingwood's notion of historical re-enactment in, for example, The Idea of History (Oxford: Clarendon Press, 1946).

simple. One of the major purposes of education is, or ought to be, to learn what is true and not merely what might be for all that is known.

I want to point out that I have no real stake in calling the learning with understanding which entails the truth of what is learned "primary" and the other sense "derivative," even though I believe the reasons I have given for doing so are quite persuasive. It is sufficient for my purposes that there are these two distinct senses and that they share as necessary conditions the all-important good reasons clause. If someone wishes to consider what I have called the "derivative" sense, "primary," I would have no serious objection. However, I shall continue to speak of learning with understanding as requiring the truth of what has been learned.

To complete this discussion let me consider in some detail one more case of learning with understanding. In this case I want what has been learned to be problematic at the present time. I think that no one would deny that Etienne Gilson has learned with understanding in some sense about Thomistic philosophy, and yet many would deny that Thomistic philosophy is true. I want to make clear just what sense (or senses) can be given to Gilson's learning with understanding. I shall take the following quotation from St. Thomas, "Therefore it is impossible that in God

His being should differ from His essence,"[15] and substitute it for p in the following schemata:

(a) X has learned with understanding that Aquinas said that p.

(b) X has learned with understanding (derivative) that p.

(c) X has learned with understanding that p.

(d) X has learned by rote that p.

Now suppose that Professor Gilson were to utter each of these sentences with the above quote substituted for p and 'I' for X (with the appropriate grammatical changes, of course). Further suppose that there are three hypothetical critics of these statements; a severe critic who not only believes that the above quotation is false, but nonsense, e.g. he believes that 'God', 'being', and 'essence', simply do not make sense, and thus could be understood by nobody; a moderate critic who, although he believes the quotation to be sensible, does not believe it to be true; and a "neutral" critic who simply doesn't have any beliefs concerning the truth or meaningfulness of the quotation, but who does believe that Professor Gilson may know what he is talking about. Then since we can distinguish judgments concerning the content of a claim from judgments concerning the justifiability of making the claim, we might get the following table of results if each critic were asked to

---

[15] Thomas Aquinas, Summa Theologica, Question 3, Article 4.

judge both the correctness of the content of the claim and the justifiability of making the claim.

		Severe Critic	Moderate Critic	Neutral Critic
I have learned with understanding that Aquinas said that p.	Claim	Yes	Yes	Yes
	Content	Yes	Yes	Yes
I have learned with understanding (derivative) that p.	Claim	Yes (?)	Yes	Yes
	Content	No	Yes	Yes (?)
I have learned with understanding that p.	Claim	No	Yes	Yes
	Content	No	No	?
I have learned by rote that p.	Claim	Yes	Yes	Yes
	Content	Yes	Yes	Yes

Thus we find the severe critic claiming that Gilson is incorrect in the case of the content of having learned with understanding that p (either in the full or derivative sense). He denies these because he believes no one could understand the terms involved in p and thus could not have <u>any</u> reason, for believing that p. For it will be recalled that in my analysis of having a reason, the reason had to be believed, and since the severe critic denies that they

can even be understood, they could not be believed, since, according to the analysis of belief, the reasons at least have to be entertained (understood in a minimal sense). He also denies that Gilson is correct in even claiming to have learned with understanding that p since such a claim would require at least that Gilson have good reasons and this the severe critic denies could happen. I have also suggested that the severe critic might allow that Gilson is justified in claiming to have learned with understanding in the derivative sense that p because he might grudgingly admit that this, being a lesser claim than full understanding could be justified by Gilson's long study and good reputation. However, he might also, if he were very critical, deny this as well, basing his denial on the grounds that the terms in p are so completely void of meaning that even the lesser claim could not be justified. The rest of the severe critic's judgments require no comment.

The moderate critic denies only the correctness of the content of Gilson's claim that he has learned with understanding in the full sense that p, and this because although he believes there to be reasons for believing p, he also believes p to be false and thus not capable of having been learned with full understanding. It is fairly obvious how he can admit the correctness of the remaining claims.

The neutral critic because of Gilson's long study and great reputation admits the justifiability of making all the claims, but simply doesn't know about the content in the case where Gilson has claimed to have learned with understanding in the full sense that p. He might also not know in the case of the derivative sense of learning with understanding about the content if he is just not sure of the meaningfulness of the terms involved.

No one, of course, would deny that Gilson had learned by rote that p nor the justifiability of his claiming this. Nor would they deny that Gilson had learned with understanding that Aquinas <u>said</u> that p, for the understanding here is of quite a different nature than the understanding connected with p itself. The former is of the sort that would include the historical nature of Aquinas' work and even the tenor of his times, while the latter would concern the system itself.

The case where X is replaced by someone other than Gilson is of no real interest here. The judgments on content would remain the same and those on the propriety of the claim would vary accordingly as to whether the name of a sixth-grader, an undergraduate in philosophy, a philosophy professor, or someone else were substituted for X.

My three hypothetical critics are, of course, not the only possible ones. There are various other combinations

of Yes's and No's for which a plausible hypothetical critic could be constructed, but it seems that the general outline is clear and it brings out rather forcefully, I believe, how the various judgments on whether or not Gilson has learned with understanding that p are related to the various beliefs of the critics and to my analyses of the various senses of learning. Thus I don't know whether Gilson has learned with understanding that p or not. That some sense can be given this claim is, I think, clear, and isolating just what sense in accordance with my analyses should help in any controversy concerning this claim.

CHAPTER III

SOME OTHER LEARNING LOCUTIONS

In the preceding two chapters, I have concentrated on the analysis of paradigm cases of "learning with understanding" and "rote learning." I have also briefly indicated the importance of this distinction and the theoretical framework it gives for understanding certain problems in education such as testing for learning and the dispositional nature of various kinds of learning; the latter throwing some light on the problem of the transfer of learning. I have nowhere indicated that I took the distinction between rote learning and learning with understanding to be exhaustive. It most certainly is not. My purpose in the preceding was rather to isolate and give a more or less complete account of two very important learning locutions. I have already indicated with my analyses of the derivative sense of learning with understanding and of the rote learning of a proposition (the 'learning that' case of rote learning) that there are a great number of learning locutions in addition to the two with which I have been primarily concerned. Some of these are closely related to the two notions which I have analyzed in detail. In this chapter I want to consider some of these other learning

locutions and certain important uses of them to see what their nature might be and how, if at all, the preceding analyses can help to explicate these other locutions and uses.

1. 'Learning How', 'Learning To', and 'Learning That'

My final anaysis of rote learning took the 'learning to' locution as its paradigm and, furthermore, related it to a specific kind of situation, S. The locution analyzed was 'X has learned by rote to q in S'. In Chapter I, I mentioned the important 'learning how' locution, but claimed that it would probably not do in an anlysis of rote learning since it was closely connected with the acquisition of skills; and some skills intuitively seem too complex to serve as examples of rote learning. In this section, I want to examine in a bit more detail this important 'learning how' locution.

The first thing to notice in this connection is that I was constrained to specify the particular situation in my expanded analysis of rote learning. Otherwise, I argued "John has learned to hold his breath" might not be a case of rote learning but rather the learning of a skill. That is, we would be more apt to say in this case that John has learned how to hold his breath, thereby giving a wider application to the learning. I think this claim can be sharpened by saying that all 'learning to' locutions can be

translated either into a 'learning how to' locution or into the paradigm 'learning to q in S' locution. Thus if John has learned to speak French, he has learned <u>how</u> to speak French. If a baby has learned to walk, he has learned <u>how</u> to walk. If, on the other hand, a baby has learned to start off on his left foot, then he has not learned <u>how</u> to start off on his left foot, but rather has learned <u>to</u> start off on his left foot whenever he begins to walk. If Sue has learned to hold her breath, she might either have learned <u>how</u> to hold her breath or learned <u>to</u> hold her breath while under water. This last case is obviously ambiguous, and we would need more information to determine which of the two alternatives is the right one.

Now I think the fact that those cases of 'learning to' which are not cases of 'learning how' are to be analyzed in terms of a specific situation is extremely important. It gives me a grip on distinguishing rote learning from the learning of skills. It also seems to point the way to a distinction <u>within</u> Ryle's 'knowing how' category; for <u>both</u> "X has learned how to q" and "X has learned to q in S" entail 'knowing how'--the first entails "X knows how to q" and the second entails "X knows how to q in S." Roughly, learning a skill is acquiring a capacity to do a variety of things in a variety of situations--the more the variety the closer is the learning of the skill to some sort of learning

with understanding; the less the variety, the closer is the skill to mere rote learning. On the other hand, rote learning is simply doing a determinate thing in a determinant situation. It depends almost not at all upon the agent's other knowledge; whereas the proper exercise of a skill depends very heavily on the agent's other knowledge. It depends on his ability to recognize the situation with which he is faced as one in which the exercise of the skill is appropriate. It involves his knowing just how to adapt his skill to difficult or new conditions. We would hardly say that someone knew how to swim if although he swam perfectly in one particular pool, he floundered helplessly in the ocean or any other pool. We would rather attempt to assimilate this odd state of affairs to rote learning by searching for the queer psychological responses which this pool and this pool alone call up in him and which enable him to swim there but nowhere else. The exercise of a skill depends on the exercise of intelligence whereas the exercise of a "skill" merely rote-learned depends only on the agent's being in the appropriate situation and intentionally exercising what he has learned.

It might be objected at this point that we can surely learn the multiplication tables by rote, but it would seem that our ability to use them in various situations indicates, according to my analysis, that we have not learned

to recite them by rote after all but rather have learned how to use them. This objection overlooks two important points which will be discussed in more detail in the following. First of all, I would claim that the capacity to use the multiplication tables in varying situations may reflect additional learning--learning that such and such are appropriate conditions in which to use the rote-learned tables. Furthermore, this additional 'learning that' may very well be 'learning with understanding that'. Secondly, there is a distinction between learning something by rote and doing something by rote. The former in a sense does not depend on our other intellectual capabilities, but the latter may very well depend on them to a large extent. Thus using the multiplication tables by rote may be an exercise of a skill which may require not only rote learning but learning with understanding as well.

The difference between learning a skill and rote learning can be gotten at in another way. As has often been noticed, skills have degrees of excellence attributable to them. We have good and bad swimmers, bridge-players, typists, mathematicians, runners, etc. There are concert pianists, piano teachers, and beginning students. There is the winning professional golfer and there is the weekend duffer. One person can have learned how to play the piano better than another and thereby knows how to

play the piano better than another. But this admission of degrees of excellence does not seem to be applicable to things which are only rote-learned. If I have learned by rote to hold my breath when my face is under water, I simply do it, and I do not do it any better or worse than anoyone else who has learned the same thing. If a child has learned to take his first step with his left foot whenever he begins to walk, he just does it, and he does not do it any better or worse than another. A boy who has learned by rote to say, "72", whenever he is asked what 8 times 9 is, simply says it, and his answer is no "better" than anyone else who has learned the same thing. When someone learns to do something by rote in a certain situation, then he either does it or fails to do it, but his performance does not have the degree of excellence associated with it that a skill has; rather it is of the all or nothing variety. Now it might be objected that some people are better at holding their breath while under water than others because they can do it longer. But upon examination this objection is seen to hang precisely on the ambiguity already mentioned in "X knows how to hold his breath." If this is taken as ascribing a skill to X, then indeed, some people can hold their breath longer than others and are better at it in all circumstances; under water, upon command, etc. But if this is taken as a piece of rote learning, then it is

simply the situation of finding one's mouth and nose covered by water and ceasing to breathe (for as long as necessary or possible). One either ceases to breathe in the situation or one doesn't, and no one who has learned to do this by rote is any better than anyone else who has learned it in the same way.

Ernest Gellner puts the point in yet another way in discussing "knowing how." He says,

> This point can also be made by saying that knowing how, like, for that matter, "knowing" tout court, is an expression for achievements. "Knowing how" is, to be precise, a triply laden expression: for when I say "Tommy knows how to skate" I (i) refer to a disposition of Tommy's to do what is describable as skating (ii) imply that he succeeds, which (iii) prejudges to some extent the question of what the criteria of good skating are, by entailing that they are such as are satisfied by Tommy's performance. Only (i) refers to a process, indeed; but important as it is to notice this, it is also important not to ignore (ii) and (iii). The sense in which "knowing how" is presupposed in all knowledge, which Ryle insists that it is in the 1945 paper, is sense (i) only. There are at least two reasons why there is a tendency to confuse the various senses: firstly it appears that we tend to be interested only in successful performances and have no terms corresponding to "knows how to. . ." but possessing only its minimal force. The nearest (axiologically) neutral term seems to be just the humble word "doing." The second reason is that some words for actions, but some only, analytically entail what the criterion of successful performance is. 'Playing football', for instance, does, though 'skating', 'composing', or 'behaving oneself' do not; the activity of playing football is defined partly in terms of trying to score goals or tries, i.e., in terms of the criteria of success.[1]

---

[1]Ernest Gellner, "Knowing How and Validity," Analysis, vol. 12.2 (1951), 26-35, 32.

At least a part of what Gellner is doing here is pointing out that 'knowing how to skate' is used to cover not only the disposition to perform certain actions, but also the fact that the "doing" must come up to a certain standard of success in order to even be described as skating at all.

The point behind all these ways of talking about 'knowing how' is that one and the same action can be described in quite a number of different ways. Thus a man sliding along the ground on his side toward a square cushion is one description of an action, but the same action can also be described as a man trying to beat out a double, avoid being tagged out, help his team score some runs, put himself in scoring position, etc. In the case of the exercise of a skill--knowing how to do something--the description of the action is in terms of the skill which in turn means that it is in terms of the minimal criteria for the action to be classed as an exercise of the skill.

An example might be helpful here. A child could sit down at a typewriter and move his fingers striking certain keys which might spell out a meaningful sentence. Suppose, by accident, the child even uses the correct fingering. We would not want to say that the child has learned how to type and therefore knows how to type, and that this was an exercise of his typing skill, even though

he performed exactly the same kind of physical movements that someone who did know how to type would have made. The point here is that although the child did something, he did not exercise a skill because his "success" was due to luck and not to the intelligent application of anything he had learned. We might possibly say that the child was "typing," being sure to use the inverted commas, but what this shows is that someone can do something in the minimal sense of "can do" that Gellner was referring to in his condition (i) without knowing how to do it, and, of course, without knowing that such and such is the way to do it. But this formulation won't quite do. The child does know how to do something (strike the keys with his fingers, perhaps), but he does not know how to type.

If knowing a skill is intimately tied up with the criteria for its successful application, then these criteria are such as not only specify the minimal physical actions which must be performed, e.g. the child's correct fingering and order of stroking, but also that these actions are performed on purpose, knowing what one is about. But if this is so, then 'knowing how to type' is not just being able to do certain things, but being able to do them in a certain way, under certain conditions, as an exercise of one's knowledge. And concomitantly learning how to type is not only learning to do things in certain situations

but also learning _that_ one does them in those situations _because_ those situations are describable in terms which make them a proper occasion for typing. And it is this _latter_ learning which accounts for the degrees of excellence associated with skills but not with 'learning by rote to'.

What I am proposing is that 'learning how' be analyzed in terms of a bunch of 'learnings to' tied together by some 'learnings that'. A person must learn to hold his breath while under water, to kick his legs and move his arms in a certain way, to breathe at certain times, to lie horizontal in the water and a host of other things in order to learn how to swim. But he must also learn to do these things in a certain way minding what he is doing. And by "minding what he is doing" I do not mean that he must be _aware_ of what he is doing in the sense of consistently telling himself what to do before or as he does it, but in the sense that he is _prepared_ to call upon his knowledge to cope with any emergencies which might arise. The problem here is to identify precisely what 'learnings that' must be included in learning how to swim. As in the case of learning astronomy in the first chapter, I do not think that this identification can in general be carried out. But it is clear that if the man _only_ learns _to_ do certain things, he learns to do them in certain specifiable

situations and could not do them if the situation varied in any great degree. I doubt whether we would say that the man has learned how and thus knows how to swim. But, of course, such a person has probably never existed who has learned only to do something in a certain determinate situation. He has also at some time in his career learned that the essential features a situation must possess in order to be an appropriate occasion for the exercise of his skill are such and such.

Schematically I am saying: 'X has learned to q' is ambiguous. If it means X has learned by rote to q in S, then the analysis I have given of rote learning is the one to be applied, and this analysis says nothing at all about X's knowledge or beliefs. Q is something X probably will simply do in S regardless of his beliefs or knowledge. It is something he could theoretically learn to do if he had no prior beliefs or knowledge at all. Of course this seldom happens in practice, and as more and more of a person's beliefs are presupposed, the more apt we are to say that he has learned how to do the thing. But this involves a broadening of the kind of situation. Situations which would not count as appropriate in the analysis of rote learning are counted as appropriate occasions for the exercise of a skill, and what I am claiming is that a person must at least have learned that these situations

are appropriate. If so-called "rote-learning" appears to have a large transfer value, it may very well actually be the learning of a skill which includes some rote learning as a part, tied together with certain 'learnings that'. It derives its justification to being called rote from its rote learned parts, and it derives its transfer value from the 'learnings that' in accordance with the ways I have indicated in the previous chapter. For if the person has not learned that certain situations are appropriate for the exercise of what he has rote learned, then he cannot be said to be acting intelligently, minding what he is doing, but only reacting to very determinate situations. In no case, however, must he be reciting to himself before he acts. One need not subscribe to the dogma of the "ghost in the machine" to insist that 'learning how' and hence 'knowing how' involve some 'learning that' and 'knowing that'.

Now it is probably evident that the cases which we ordinarily call rote learning are not such as presuppose absolutely no prior beliefs or knowledge. They almost always do, but what I am arguing is that they need not. In the classic learning the multiplication tables by rote example, it is almost invariably the case that the student understands the English language, has some beliefs about the meanings of many of its words, and recognizes in some

more or less vague way the function and use of arithmetic. All of these things tend to make his rote learning of the multiplication tables easier and especially tend to dictate the type of teaching methods used, e.g. a verbal or written presentation of the tables and drill. What I am claiming is that although this prior knowledge is made use of, it need not be. The boy (or a parrot) could have been taught, a la Skinner perhaps, simply to recite the tables upon the presentation of a suitable cue without having a clue as to what he was about except maybe that he was rewarded for the performance or escaped punishment for non-performance. The ordinary cases of rote learning derive their title to being "rote" because although they may have made use of prior knowledge, they are such that the prior knowledge has helped only in making the learning easier. What has been learned is not in a real sense an addition to the pupil's knowledge. He might as well, for all the good the rote learning does him, have learned it without making use of the prior knowledge. The sole increase in his "intellectual" capacity is in being able to give a certain limited performance in certain highly determinate situations. As Ryle would put it, the acquired capacity is not a quality of mind or intelligence. It is not something that the agent does intelligently or minding what he is doing. Nevertheless we do say that he knows how to do

what he has learned by rote. What I am arguing is that this 'knowing how' is 'knowing how' in a very minimal sense and it is to be sharply distinguished from 'knowing how' which involves intelligence.

Let me consider another example; this time one that Ryle proposes. He says,

> But it would be quite possible for a boy to learn chess without ever hearing or reading the rules at all. By watching the moves made by others and by noticing which of his own moves were conceded and which were rejected, he could pick up the art of playing correctly while still quite unable to propound the regulations in terms of which 'correct' and 'incorrect' are defined. We all learned the rules of hunt-the-thimble and hide-and-seek and the elementary rules of grammar and logic in this way. We learn how by pracrice, schooled indeed by criticism and example, but often quite unaided by any lessons in the theory.[2]

I think that this example is also open to the criticisms which I made of the swimming case. Indeed, the boy learns to move the pieces in just those ways which are sanctioned by the rules. However, he must also, at some time, have learned that these moves are to be made in the situation describable as playing chess, against an opponent, and not just moving them around in accordance with the rules whenever he feels like it. In addition to this sort of 'learning that', however, there are in this case, as in most complex examples of 'knowing how', even more examples of 'knowings that' which must be satisfied before we are

---

[2]Ryle, op. cit., p. 41.

willing to say that the boy has learned how and hence knows how to play chess. What Ryle has described is at most someone learning to move the pieces in accordance with the rules. This I would admit could be learned by rote. But there is more to chess than refraining from making illegal moves. One must also learn the point of the game, and to do this it seems one must know that the point of games in general is that there is a goal or end-position to be striven for by both the players, and this end-position constitutes winning, and given that one plays games at all, winning is a desired outcome. A student could learn to move the pieces correctly as Ryle suggests and still not have learned how to play chess simply because he doesn't know that the object of the game is to force his opponent's king into a certain kind of situation. Even if it were possible to learn by rote to force one's opponent into position p_1 or p_2 or. . ., this is not the way anyone learns how to play chess. Rather we learn that the object is to force one's opponent into a certain kind of situation however or wherever it occurs on the board. We must know that this kind of situation is what constitutes "winning" at chess and that "winning" is the point of the game even though we may never recite these pieces of knowledge to ourselves or anyone else.

In the above example and the ones that follow,

Ryle is trying to drive a wedge between knowing in the sense of having acquired a capacity to do something and knowing in the sense of being able to recite certain facts and prescriptions to oneself or others. This is for Ryle an attack from one direction on the dualism between mind and body, for dualistic theories often seem to be claiming that some sort of shadowy intellectual recitation must occur before any action can be called intelligent. To this end Ryle distinguishes sharply between 'knowing that' and 'knowing how', claiming that the latter cannot be "reduced" to the former and thus that certain exercises of intelligence or knowledge need not have any connection with the "intellectual" kind of knowledge usually expressed by 'knowing that'.

Now, what does the line of argument I have been pursuing do to Ryle's distinction between 'knowing that' and 'knowing how'? What does it do to his arguments against the dual-world dogma? Well, it leaves entirely untouched his major argument against the ghost in the machine. Ryle is right in saying that we do not have to preach before we can practice. My arguments also leave untouched Ryle's major distinction between 'knowing how' and 'knowing that', We know propositions in an all or nothing sort of way and skills in a degrees of excellence sort of way. Nor can we "reduce" in many cases 'knowing how' to 'knowing that'.

But granting Ryle these points, it seems entirely gratuitous of him to claim as he does, that 'knowing how' entails no 'knowing that' statements. He would have to make this claim only if 'knowing that' necesssarily involved some sort of preaching to oneself. But this is false on grounds entirely independent of the 'knowing that'--'knowing how' distinction. A person can know that the way in which one wins at chess is by forcing one's opponent into a certain kind of situation without ever having "recited" this to himself or anyone else. As far as 'knowing how' itself is concerned, I have argued that it is ambiguous. One can know how to do something in the sense of having acquired a skill. In this case 'knowing how' entails not only simply being able to do things in certain situations, but being able to do them on purpose, knowing their point and relation to the criteria of success for the skill. On the other hand, 'knowing how' also has the minimal sense of merely being able to do certain things in certain highly determinate situations without having to know the point or criteria of success of the doing. In addition this latter case of 'knowing how' has no degrees of excellence associated with it because it has no connection to any criteria which the agent must know, and these criteria, which are present in the former case, are such that the agent must know that what he is doing is an attempt to satisfy them.

Actually Ryle seems to be dimly aware of this ambiguity in 'knowing how', although he is loath to bring it out in the open for he seems to fear unnecessarily that it would damage his arguments against the dual-world theory. He says:

> To say that a sugar-lump is dissolving, a bird migrating, or a man blinking does not imply that the sugar has learned to go liquid, that the bird has learned to fly south in the autumn, or that the man has learned to blink when startled. But to say that a soldier obediently fixed his bayonet, or fixed it in order to defend himself, does imply that he has learned some lessons and not forgotten them. The new recruit, on hearing the order to fix bayonet, or on seeing an enemy soldier approaching, does not know what to do with his bayonet, how to do it, or when to do it and when not to do it. He may not even know how to construe or obey orders.[3]

The lessons that the soldier has learned and not forgotten seem to be <u>that</u> in order to fix his bayonet obediently, he must intend to fix his bayonet when he hears the order. Nor is Ryle completely willing to say that the soldier merely learned by rote <u>to</u> fix his bayonet by reacting to the verbal cue, for that would make the soldier an automaton and not a person intelligently exercising his skill of knowing how to fix his bayonet. That this is so is shown as Ryle goes on to say

> Children, semi-lietrates, old-fashioned soldiers and some pedagogues tend to suppose that being taught and trained consist in becoming able merely to echo the exact lessons taught. But this is an

---

[3] <u>Ibid</u>., p. 146.

> error. We should not say that the child had done more than begin to learn his multiplication-tables if all he could do were to go through them correctly from beginning to end. He has not learned them properly unless he can promptly give the right answer to any snap multiplication problem (lower than 12 x 13), and unless he can apply his tables by telling us, e.g. how many toes there are in a room in which there are six people. Nor is a man a trained rock-climber who can cope only with the same nursery-climbs over which he was taught, in conditions just like those in which he was taught, and then only by going through the very motions which he had been then made to perform. Learning is becoming capable of doing some correct or suitable thing in _any_ situation of certain general sorts. It is becoming prepared for _variable_ calls within certain ranges.[4]

Ryle is here describing someone learning _how_ to do something--acquiring a skill--and not just learning by rote _to_ do certain things in certain highly determinate situations. When he uses the phrase, "learned them properly," in speaking of the child's learning of the multiplication tables, he seems to mean that "learning properly" is 'learning how' and "learning improperly" is 'learning to'. For, as he admits, some learning has gone on even if the child can only go through the tables--rote learning--but it is not "proper" learning because the child has not yet learned _that_ the multiplication tables are useful in a variety of situations and has not learned _that_ these situations are related to his rote learning in certain ways. When he acquires this latter learning, then he will have learned the

---

[4] _Ibid._, pp. 146-147.

multiplication tables properly; for then he will have learned how to multiply.

Where does this leave me with regard to the relations among 'learning that', 'learning how', and 'learning to'? Well, with most common cases of learning, the foregoing shows that these are intimately related. We cannot in general learn any skill (learning how) or anything which requires some sort of skill without learning by rote to do certain things in certain situations. However, we must also learn that certain related situations are describable in ways which make the performance of the learned responses appropriate. Both 'learning to' and 'learning that' are involved in 'learning how'. Now some of the 'learnings that' involved in 'learning how' will be 'learning with understanding' and some not, and one way in which the degree of excellence which a person has in a skill can be increased is by increasing the amount of learning with understanding.

Of course, another way to increase the level of excellence is to increase the efficiency of the subskills, e.g. shorten the time between the situation and the action. In this connection I am supposing that it often occurs that an apprentice in a skill must stop when faced with a new situation and go through a conscious process of judging what the situation is like so that he can tell what piece

of 'learning to' is appropriate. He consciously utilizes his 'knowledge that' in deciding which of the things he has learned to do applies. As he becomes more skilled, this conscious judgment time grows shorter and shorter and eventually vanishes. It does not become instantaneous as Ryle has pointed out. Rather the apprentice becoming a master is learning to act automatically without thinking at all what he is doing. It must be noticed, however, that this does not entail that he no longer 'knows that'. Rather his 'knowing that' has become the basis for a new 'learning to'. Thus the apprentice carpenter may have to stop and think what to do when he encounters a wet board, but the master carpenter has learned to do whatever is necessary without thinking. Both probably know that such and such is the proper way to treat the board, although neither of them need even be able to say with any precision what this way is aloud or to themselves. (It is important to recall in this regard that although 'knowledge that' or 'belief that' are most often ascribed on the basis of certain verbal behavior, they can be ascribed on the basis of non-verbal behavior as well.) As Broudy says

> In other words, if we keep the theoretical level constant, differences in mastery are differences in skill; if we keep efficiency of response constant, differences in mastery correspond to differences in the levels of theoretical insight required to get the correct response.[5]

---

[5]Harry S. Broudy, "Mastery," in Language and Concepts in Education, ed. by B.O. Smith and R.H. Ennis (Chicago: Rand McNally & Co., 1961), p. 84.

I think the foregoing points out another very important sense of 'learning to' in addition to learning by rote to q in S. The master carpenter has learned to q in S although he has not learned to do it solely as a result of his experience in situations of kind S, but also as a result of his knowledge. This 'learning to' locution shares with rote learning the necessary condition that the probability that X would q if he were now in S is nearly one, and probably the condition that X be in a series of S's as well. It is not, however, necessary that the trials alone produce the learning. This case is to be distinguished from the earlier case in which I claimed prior knowledge only helped in making the learning easier or determined the teaching methods to be used. In that case what was learned was only an habitual response which, lacking any widespread applicability to different situations because of its relative independence of the agent's other knowledge, was in a real sense <u>not</u> an increase in the agent's intellectual capacity. However, the earlier case does share with the present one the virtue of the response's being more efficient by being made habitual. The present case differs, however, from the earlier one in that the master carpenter knows what he is about even though the response has become routinized. He <u>can</u>, if necessary, detect lapses into bad habits in the exercise of the

routine and correct them, whereas the merely rote learner cannot. The master's 'learning to' is actually a 'learning how to routinely'--without thinking--although he can apply his knowledge to it if necessary. He has not in one sense increased his intellectual capacity, but he has increased the degree of excellence in his skill by routinizing certain parts of it thereby freeing his mind for other tasks connected with his carpentering.

We might say of the master carpenter that he has learned to q by rote, but this is something different from learning by rote to q. In the former case the phrase "by rote" qualifies the doing, not the learning, and is used as a term of approbation. I have already pointed out that this locution shares the habitual nature (conditions [1] and [2]) which I have associated with rote learning but differs from it in being essentially based on prior knowledge. Doing things by rote is to be commended whenever the doing is the exercise of a knowledge how which is always ready to correct the doing. In the latter case, on the other hand, "by rote" qualifies the learning, and not the doing, and is used as a disapprobative term since it implies that the learning is not essentially tied to any prior knowledge and could not be corrected by such knowledge if things began to go badly.

Two examples will help to illustrate the points I

have been making concerning the ambiguity in 'knowing how' and the distinction between rote learning and rote performance. First, consider the case of a normal child who has just recently learned to walk. We say of him that he knows how to walk. This means, I think, that he has simply learned by rote to walk and can now, on reasonably level ground, walk. This is the minimal sense of 'knowing how' which does not involve a very wide applicability nor the child's knowing that what he is doing is an attempt to walk. He simply does what he does. However, the child very rapidly learns other things connected with walking, how to turn around, how to climb stairs, how to walk on uneven ground, etc. He begins to grasp the criteria of success in walking and when walking is appropriate. He learns that, and possibly knows that, one walks in certain situations for certain purposes. He now knows how to walk in the sense of having acquired a skill with fairly wide applicability and he exercises this skill on purpose, minding what he is doing. Very probably in these early stages he must pay attention to his walking and concentrate on making the right moves at the right times. Later he learns to walk in ordinary circumstances by rote. His <u>performance</u> is now rote although his prior learning was not totally rote. Later, as a result of some injury perhaps, he may have to go through this process again. Analogously,

knowing how to walk in ordinary circumstances and doing so by rote does not mean that one knows how to walk in special circumstances, e.g. climbing mountains, or walking on a bed of nails. Further learning is necessary in these special cases and one probably has to again concentrate one's attention on the task. One can no longer simply perform by rote.

The second example concerns people such as coaches and teachers. It seems that we would want to say of these people that they know how to do certain things even if they are incapable of performing the actions themselves. The first point to be made is that although Ryle's arguments against the reducibility of 'knowing how' to "knowing that' are correct, this does not mean that there are no cases where 'knowing how' is not to be explained in terms of 'knowing that.' His argument simply shows that 'knowing how' need not be explained in terms of knowing that such and such is the appropriate way to proceed. I think the present example provides cases where 'knowing how' is to be understood in terms of 'knowing that', plus, perhaps, a certain pedagogical skill in imparting this information to one's students. In addition, the successful coach often could, at some previous time, have actually performed the appropriate actions, but time and old age now have taken their toll. In these cases, although the actual skill is

absent, the knowledge of appropriate circumstances for the exercise of the skill and the knowledge that such and such is the way to go about performing the skill together with the ability successfully to train others in the skill surely justifies us, at least by courtesy, in saying that the coach or teacher 'knows how' to do certain things.

Finally 'learning that' whether it is with understanding or not, generally involves some previous or concomitant 'learning how'. In general a person cannot learn that the Declaraction of Independence was signed in 1776 unless he has learned how to understand the language in which that proposition is expressed to him, orally or in writing. This having learned how to understand a language seems in general to be a minimal requirement for 'learning that'. In particular cases, other 'learnings how' and 'learnings to' may also be present, as will be shown in more detail in the following section.

2. Learning a Subject Matter or Discipline

So far I have been concerned only with learning locutions which ordinarily take rather limited propositions or responses as substitution instances. In this section I will consider how these "particular" locutions are related to the learning of a whole body of material which could constitute what might be called a "subject matter" or discipline." For example I shall consider such locutions

as, "He has learned some elementary physics," "John has learned the basic principles of logic," "She is learning European history," etc.

As a first attempt, it might be suggested that to learn a subject or discipline one needs only to learn all, or most, or a certain important subset, of the facts of that discipline. This would reduce learning a subject to the form, learning <u>that</u> p_1 & p_2 & . . . where at least <u>most</u> of the p_i's are true. I have already argued in Chapter I that to learn a subject requires that a great number of the propositions that are learned must be true, although it may be impossible to specify some minimal subset, <u>all</u> of whose members must be true and have been learned in order to say that one has learned the subject. Following this line a person has learned a subject with understanding just in case a large number of the p_i's that have been learned have been learned with understanding. It is important to notice here that the propositions, the p_i's, may have many different forms. Some of them may define the technical terms of the subject, some may give the basic laws or hypotheses, others may report particular facts about the subject, still others may give some relations between particular facts or between particular facts and laws.

Although the broad range of the p_i's makes this view appealing, it is, I think, over-simplified. For one

thing even in order to learn that p_1, a person must minimally learn or have learned how to understand the lessons he is given, either orally or in writing. This is the minimal requirement I mentioned at the end of the last section. But of even more importance, there are certain skills and techniques associated with almost every particular subject which any student of the subject must learn how to apply. It is logically possible that a student could have learned that p_1 & p_2 & . . . and yet not know how to apply his knowledge to new situations. It is unlikely, to be sure, but even in the case of having learned with <u>understanding</u> that p, it is just possible that the student has learned with understanding that All B's are C's (a law of the subject), and learned with understanding that b is a B and yet fail to draw the inference that b is a C. Similar examples could be constructed for other forms of inference, e.g. inductive.

The problem here is that the student has failed to draw the inference simply because he lacked the skill which comes from practice in drawing inferences of this kind. He has not learned <u>how</u>, and thus does not know <u>how</u> to put his 'knowledge that' to use. Of course, it could happen that even if the student has learned how to draw this sort of inference, he may not do so in particular cases owing to some sort of psychological interference or

block. However, the point that I am making is a conceptual one. If the student has not learned how to do certain things as well as learned that certain things are the case, then we might not want to say that he had learned the subject matter, although this would probably depend on just how important the skills were in the particular subject. In arithmetic, for example, it seems highly unlikely that a student who has learned that 1+1=2, 1+2=3, 1+3=4, . . . would be judged to have learned arithmetic if he could not apply his knowledge to calculating sums of large numbers, i.e. if he had not learned how to add large numbers.

Another example might be helpful. A student may have learned that poets frequently use metaphors to express a thought, and he may have learned that a certain sentence in a certain poem is a metaphor. But unless he has also learned how to interpret metaphors in poetry, he will have no idea what the poet is trying to express. And this learning how is usually best accomplished through practice in interpreting poems.

Thus the 'learning that p_1 & p_2 & . . .' model tends to overemphasize sheer memory of facts either by rote or with understanding without sufficient attention to the acquiring of the necessary skills and techniques. Learning a subject matter usually involves all three learning locutions, 'learning that', 'learning how', and 'learning to'.

Learning a subject matter with understanding involves both many of the 'learnings that' being 'learning with understanding that' and increasing the level of excellence of the 'learning how'. The latter, as I have already argued in the last section involves yet more 'learnings with understanding that', and routinizing certain subskills.

Of course, the dangers of not appreciating the mixture of learnings which go into learning a subject matter have long been recognized. Simple factual recall tests of the kind so often found in history courses which stress the memorization of historical detail often fail to give due weight to an understanding of the causes and consequences of historical events. Further, they do not test a student's skill in analyzing historical phenomena. Tests of arithmetical skills often fail to test whether the student has acquired the understanding necessary to apply his skills to practical problems. Even tests which purport to test the basic understanding of something like economics sometimes lead to glib generalizations which are not supported by basic skills and facts. For example, a glib exposition on the "laws of supply and demand" may not test the student's ability to translate statistical information into the terms of supply and demand, and thus his ability to test his generalizations. And, of course, there are permutations and combinations of all these dangers. What

is required is that the teacher be fully aware of just what types of learning he is after and the relative weights to be placed on each. He must then communicate this to the students and seek tests which are good clues to the desired learnings. The use of the three learning locutions--'learning that', 'learning how', and 'learning to'--and an awareness of what they mean and how they are related in any given subject should aid immeasurably in curriculum planning, the selection of teaching methods and materials, and test construction. It has been one of the purposes of the foregoing analyses to begin to point out the complex relations of these types. There may even be others suggested by my analyses which I have not mentioned, but the three I have discussed seem to be especially important.

3. The Learning of Evaluative Notions[6]

In this section I want to consider the effects of substituting for q in my three learning locutions--learning that q, learning how to q, and learning to q--expressions containing evaluative terms. For example I want to consider the relations and differences between the following:

A. X has learned that one ought to be honest.

B. X has learned how to be honest.

C. X has learned to be honest.

---

[6]I am especially indebted in this section and the following one to Israel Scheffler's treatment of these and related issues. Israel Scheffler, The Language of Education (Springfield, Ill.: Charles C. Thomas, 1960), especially Chapters 4 and 5.

Scheffler in the above mentioned book notes an ambiguity in sentences like A.--'learning that' locutions where what follows the 'that' is what he calls a "norm-stating" sentence. This ambiguity is between an "active" and a "non-active" interpretation of A. In the active interpretation of A. we suppose that the student has acquired as a part of his character the norm of behaving honestly. He has learned to be honest. He returns excess change given him by careless cashiers, he doesn't cheat on his income tax return, he carefully handles other peoples' money, etc. He is an honest man. In the non-active interpretation of A., X need not have learned to be honest, nor are any examples of his being dishonest sufficient to disprove A. Rather the student gives verbal assent to A. and this is all that is required. The ambiguity is meant to mark the notorious fact that people sometimes deliberately act contrary to what they know is right. They sin knowing that they sin.

Now although I think Scheffler is on the right track with this distinction, I believe he oversimplifies it tremendously. For example he says that any case of dishonesty by a person would refute the ascription of A. in its active interpretation. This, however, simply won't do. If, as Scheffler seems to claim, A. in its active interpretation is the same as C. (or at least entails C.),

then isolated examples of dishonesty will not disprove it. There are two reasons for this. To begin with I don't think Scheffler would object to my claiming that A. in its active sense entails (i) that X believes that one ought to be honest and furthermore, (ii) that X has learned to be honest, which in turn entails at least that the probability is nearly one that X would be honest if he were now faced with the choice of doing an honest or a dishonest thing. (Cf. my remarks on 'learning to' in the preceding.) Now I have already argued that while belief is dispositional and thus anyone who has a certain belief would do things in a certain range if he were in certain situations, I also argued that there were very few, if any, <u>known</u> laws supporting these dispositional statements. And lacking these laws, it always seems possible to find an explanation of the absence of the anticipated behavior consistent with the belief. Thus a person may not have acted honestly in a given situation even though he has learned in the active interpretation that one ought to be honest simply because he did not believe the situation was one in which the norm of honesty was applicable. Or he may have failed to act honestly because although he believed the norm of honesty was applicable in the situation, it was overridden by other considerations. All that is required is that the person would <u>generally</u> do the honest thing.

The second reason why isolated examples of dishonesty will not refute the active interpretation of A. is because the probability of a person's being honest is only approximately equal to one; it is not equal to one. A man who only rarely does a dishonest thing because of some temporary weakness of the will is not thereby judged to be dishonest. Men are imperfect and isolated sins do not make an evil man. The reason for this seems to be the fact that 'honesty' can be applied either to the characters of men or to specific acts although the former seems to be the primary use; for the honest act is generally thought of as what an honest man would ordinarily do. What I am arguing is that the honest man may perform an occasional dishonest act without thereby being judged to be dishonest. Of course, I admit that a man would not be called honest if he very often performs dishonest acts although the dividing line is hard to draw; for it always seems logically possible to construct a situation in which the man performs acts describable by others as dishonest because he always felt that there were overriding considerations. Of course, this seldom occurs in practice. However, with a virtue such as courage it would be easy to imagine a man who had learned that one ought to be courageous (active interpretation) and yet in the one situation in which he might have been courageous, he either didn't believe it to be a situation

in which courage was applicable, or he thought the reasons for being courageous were overridden by other considerations, or he had a momentary weakness of will. All that seems to be required is that he would have been courageous in almost all other situations calling for courage, although he simply was never faced with any. In fact this sort of situation is the basis for many a story plot. The hero fails to manifest a virtue in a particular situation for one of the reasons given above (usually because there are overriding considerations), the community judges (wrongly) that he does not have this virtue, and the rest of the plot is concerned with showing that he actually does have the virtue by exhibiting other situations in which he does display the virtue in question. The problem in real life is that often the "hero" hasn't got a good writer to make sure he has the opportunity later on to display the virtue which he seems to lack.

Likewise with the non-active interpretation of A. Scheffler oversimplifies the case. Having learned that one ought to be honest entails believing that one ought to be honest in the non-active as well as the active case. And as I have already argued not only does believing a proposition involve certain verbal behavior in certain situations, it also involves doing certain things. Thus Scheffler must be wrong in claiming that no examples of

dishonesty are sufficient to disprove the non-active ascription of A. The man who consistently behaves dishonestly while piously proclaiming that one ought to be honest would not be said to have learned that one ought to be honest even in the non-active interpretation. Hypocrites are, unfortunately, always among us.

What Scheffler probably has in mind in his talk of the non-active interpretation of A. must be something like

A'. X has learned that society <u>says</u> that one ought to be honest, or

A''. X has learned that most people <u>believe</u> that one ought to be honest.

Now either A'. or A''. or others like them surely <u>are</u> consistent with a person's continually behaving in a dishonest way, although as I have just argued, A. itself is not. This is even more obvious if A. is learning with understanding in the non-active sense. It would be <u>extremely</u> odd to say that someone has learned with understanding that one ought to be honest (non-active) and yet claim that no examples of dishonesty could refute this, although they could not refute A'. or A''. if <u>they</u> were learned with understanding.

I have been arguing that although Scheffler's distinction between an active and a non-active interpretation of 'X has learned that one ought to be honest' is appealing, it is grossly oversimplified and even misleading in that

it leans far too heavily on a strictly behavioristic interpretation in one case (active) and far too heavily upon an extremely intellectual interpretation in the other case (non-active).

Let me try to recast Scheffler's distinction between the active and non-active interpretation of "X has learned that one ought to be honest." In the active interpretation this entails 'X believes that one ought to be honest' and 'X has learned to be honest'. In accordance with the remarks I have already made, 'having learned to be honest' involves the probability of X's performing honest actions being nearly one. Thus a fairly large number of dishonest actions on someone's part would tend to refute the ascription of "He has learned that one ought to be honest" in its active interpretation. I have argued, however, that isolated cases of dishonesty would not refute this ascription and even that it is logically possible that a great number of cases might not refute it depending on the further circumstances.

The non-active interpretation of "X has learned that one ought to be honest" entails only that X believes that one ought to be honest. He need not have learned to be honest as well. In this case acts of dishonesty by a person are much less important to the truth of "He has learned that one ought to be honest" than in the active interpretation.

However, I have argued that they are not totally irrelevant. The difficulty arises with the case of a very consistently dishonest man who nevertheless maintained that one ought to be honest. Since believing as a disposition involves doing as well as saying, we could ask if the man really believes that one ought to be honest. We might, as Scheffler seems to urge, maintain that the man has learned that one ought to be honest (non-active), even though he consistently acts dishonestly; or we might say he has not learned, even non-actively, that one ought to be honest, but rather he has learned that society says that one ought to be honest or that most people believe that one ought to be honest. It is the possibility of this latter case which prompts me to say that Scheffler's original way of distinguishing the two cases is misleading. To repeat, the difference between the active and the non-active interpretation of "X has learned that one ought to be honest" is that the former entails X's having learned to be honest while the latter does not. The non-active interpretation of A. (as opposed to the A'. or A''. interpretation) probably fits best those cases where the failure to act honestly in a great number of cases is due almost entirely to a rather prevalent weakness of the will. The man would have been honest except for this character flaw. However, consistent dishonesty may lead us to question a mere

weakness of will and tend to make us believe that A'. or A''. is the proper description rather than A. in its non-active interpretation.

Thus Scheffler's non-active interpretation of 'X has learned that one ought to be honest' is ambiguous. It may mean either (a) the person has not learned that one ought to be honest but has rather learned something like 'Most people <u>believe</u> that one ought to be honest', or, (b) the person has learned that one ought to be honest, but because of a prevalent weakness of will, he seldom acts honestly. He <u>would</u> act honestly except for this psychological block.

Scheffler does, however, correctly point out the dangers in failing to make the distinction. We might have succeeded in establishing that a student has learned that one ought to be honest in the non-active interpretation, e.g. by his successful performance on certain verbal-type tests, and then claim that it followed necessarily that the student has learned that one ought to be honest in the active interpretation. That is, we might claim to have necessarily trained a student who <u>was</u> honest--one who has learned to <u>be</u> honest--when in fact we had not. The danger is apparent. In addition, failing to make the distinction is greatly facilitated by the extreme difficulty of actually finding out if a student has learned that one ought to be

honest (active interpretation). For we would have to observe his conduct over a long period of time in situations in which the norm of honesty is presumed to be applicable; whereas giving him a verbal test is so easy.

However, there are dangers inherent in Scheffler's original distinction as well. For if we believe that a single dishonest act refutes the ascription of "He has learned that one ought to be honest" in its active interpretation, we may do the student a great injustice. He may have learned to be honest and yet commit occasional dishonest acts for reasons which I have already discussed. There is also a danger in saying that dishonest acts have no relevance to whether a student has learned that one ought to be honest in the non-active interpretation. This danger is especially brought out if the learning is learning with understanding. For even if "One ought to be honest" is neither true nor false, reasons can be given for it. It is not something which can be merely intuited by anyone.[7]

---

[7]This is the problem of the "supervenience" of evaluative notions. Very briefly the problem is this: We very often cannot give a list of factual (true or false) statements which are equivalent to calling something "good," for example. Nevertheless the factual properties of a thing are connected with the evaluative judgments we make concerning it. For example, it would be a mistake to call one baseball bat "better" than another if both had exactly the same properties except for spatio-temporal location. There must be a reason for the one's being better than the other.

But if this is so, then it seems that a student could not have learned with understanding that one ought to be honest even in the non-active sense unless he has reasons for this belief. Furthermore, these reasons are for a belief which is general in that it applies to everyone. That is, "One ought to be honest" applies to the student as well as everyone else unless there are special circumstances which themselves are general, i.e. exceptions may be made or the norm overridden in special circumstances, but the logic of evaluative notions requires that the same exceptions be made for anyone if the circumstances are the same.[8] Now in the light of my previous analyses of learning with understanding and having reasons for a belief, the consistently dishonest student might very well be said not to have learned that one ought to be honest even non-actively, but rather learned that people say one ought to be honest or some such thing. Whether the connection between evaluative beliefs and action be contingent or not, it is very important, and consistent action contrary to that specified by a supposedly "believed" norm may very well lead us to say that the norm is not believed at all. The danger then is in not recognizing this close connection between

---

[8]See R.M. Hare, The Language of Morals (Oxford: Oxford at the Clarendon Press, 1961), and Freedom and Reason (Oxford: Oxford University Press, 1963), for a discussion of this generality requirement and a more detailed account of the "supervenience" of evaluative notions in addition to a treatment of other issues related to the present work.

evaluative beliefs and the sort of actions specified by them. We might be tempted to say of a student who consistently acts dishonestly that he has learned with understanding that one ought to be honest (in the non-active interpretation), thereby implying that there is nothing wrong with his cognitive grasp of evaluative notions, but that he only needs to be taught to _be_ honest or only needs help in overcoming his very pervasive weakness of will. But it might be _just_ the lack of a proper cognitive grasp of the evaluative notions which leads him to act dishonestly, i.e. he may believe that "One ought to be honest" _means_ only that people _say_ one ought to be honest and thus not recognize the imperative force of the norm on him.

Although learning that one ought to be honest entails in the active interpretation that one also learns _to_ be honest, it is possible to learn to be honest without learning that one ought to be honest, without believing that one ought to be honest, and even disbelieving that one ought to be honest. Learning to be honest can be done by _rote_; for all that it entails is that certain patterns of actions probably occur in certain cituations. We are all too familiar with the person who merely behaves in accordance with customary morality simply because he has learned _to_ do so and not because he has learned that he _ought_ to. Faced with difficult moral choices or evaluations such a

person is at a loss as to what to do; for he has not learned how to deal with any but the stereotyped situations.

In accordance with my previous analyses learning how to be honest includes roughly learning to do certain things in certain situations and learning that these things are to be done in situations describable in certain ways which make the norm of honesty applicable; and learning that the essential characteristics of such situations are of such and such a sort so that new occasions for being honest can be recognized. For example, in order to have learned how to be honest, a person may have learned to count his change when he pays a bill and learned that whenever he has received too much change, it is honest to return the excess. He may also have learned that a situation in which he receives more of anything than his due (perhaps because of a machine malfunction he gets a salary check for more than his salary) is sufficiently like the foregoing for the norm of honesty to be applicable.

Thus although learning how to be honest involves certain 'learnings to' and other 'learnings that' tying the 'learnings to' together, it does not entail learning to be honest nor learning that one ought to be honest. That learning how to be honest does not entail learning that one ought to be honest, is easy to see. One can learn to do the required things and learn that these

things done in certain situations are called "honesty." One can learn the essential characteristics of these situations without learning at all that one ought to be honest, perhaps without understanding what 'ought' means. It is also clear that a person could learn how to be honest without learning to be honest. The reason for this is very simply that, as Ryle would put it, 'learning how' is a capacity verb while 'learning to' is a tendency verb. One can acquire the capacity for being honest without tending to be honest. Furthermore, in this case I have argued that the tendency to be honest (performing actions describable as honest) does not imply that one has the capacity for being honest. (The person may lack the intellectual background for intending to perform actions describable as honest.) This is just one more example of the ambiguity in Ryle's 'knowing how' category. One can learn to be honest and thus know how to be honest in the minimal sense of being able to perform on purpose certain actions in certain situations without having learned how to be honest and thus not knowing how to be honest in the full sense of 'knowing how' which involves not only being able to do certain things, but being able to do them with the intention of being honest, minding what one is doing.

Keeping in mind the foregoing analyses, the central problem of moral education can now be formulated in

terms of the three learning locutions with which I began this section. The honest man is the man who has learned to be honest in the way he has learned how to be honest because he has learned that one ought to be honest. The dangers in moral education arise from not fully appreciating the independence of all three types of learning; for it is all too easy to satisfy oneself that one or another of the learnings has been accomplished and then jump to the conclusion that the student is thereby an honest man.

4. Learning and Teaching

Scheffler gives an excellent analysis of 'teaching' in the above-mentioned work. I do not believe that I can materially improve on his account; however, I believe that it would be worthwhile to go over his main points defending them against possible objections and emphasizing the relations between teaching and learning. This will be the purpose of the present section.

The first point that Scheffler makes is to distinguish between two possible senses of 'teaching'. These are the "success" and "intentional" uses of the verb. As Ryle, among other philosophers, has pointed out, verbs describing activities (intentional) are generally to be sharply distinguished from verbs describing "the appropriate upshot" of those activities (success).[9] Thus 'hunt',

---

[9]Ryle, op. cit.

'look for', 'listen', etc. all describe activities in which people engage. The appropriate upshots of these activites are respectively, 'bag', 'find', 'hear'. One can hunt lions where there are no lions, but one cannot bag a lion unless there is a lion to be bagged. A person can look for a needle in a haystack without the needle's being there, but he could not find it unless it were there. A man could listen for a burglar who was not there, but could not hear him unless he were. (Of course, he could hear noises which he <u>believed</u> were made by a burglar, but he could not hear a burglar's noises unless there <u>were</u> a burglar.)

The problem with 'teaching' is that it is used <u>both</u> as an intentional verb describing an activity <u>and</u> as a success verb describing the appropriate upshot of this activity. This, of course, leads to confusion if these senses are not kept distinct. The appropriate upshot of teaching as an activity is, of course, learning. Thus, paradoxical as it may sound, it would be true to say that I have been teaching Jones logic, but I have not taught him any. The paradox vanishes when it is noted that the first use of the verb 'to teach' is an intentional one--I have been engaged in trying to get Jones to learn logic--whereas the second use is a success use--Jones has failed to learn logic despite my efforts.

Scheffler uses this distinction to completely

destroy the controversy over whether there can be teaching without learning or not by showing it to be a mere verbal confusion. Those who urge that there can be no teaching without learning are using 'teaching' in its success sense and what they say is analytically true. On the other hand those who urge that there can be teaching without learning are using 'teaching' in its intentional sense and what they say is analytically true also. The controversy cannot be taken literally for it dissolves as soon as one or the other of the two uses of 'teaching' is specified. Rather the advocates of "no teaching without learning" are urging us to concentrate on the learner, to evaluate our teaching methods at least in part by the learning they produce. On the other hand, those who cry "there can be teaching without learning" are urging administrators and teachers not to evaluate teaching solely on the basis of the learning produced because even the best teachers may fail to produce learning in absolute dullards while even the poorest teacher may contribute to the learning of exceptionally bright students.

This latter point is connected to another facet of "success" words. If one can be successful in teaching, one can also fail to achieve the desired results; learning simply may not occur. Nevertheless, teaching always involves trying to get the student to learn what is being

taught. 'Teaching' in its intentional sense is engaging in certain activities intending that these activities produce the learning in the student of what is being taught. Of course, the "trying" may not be limited to any particular lesson, and the goal envisaged may lie beyond the bounds of the whole teaching activity. I may be teaching my son how to play baseball by playing catch with him although I may not be trying to teach him (successfully) in the bounds of any given lesson nor even in the bounds of all the lessons I give him. He may have to practice after I quit teaching him and even have other teachers, e.g. coaches, before he eventually learns how to play baseball. Nevertheless I have been teaching him how to play baseball since I have engaged in activities with him, the intention of which was to contribute to his learning how to play baseball.

Thus teaching involves trying to achieve an intended result, and failure is always possible despite the best efforts of the teacher. The student or the universe may not cooperate. The problem here is that there is no way to clearly formulate the criteria for when a teacher has done the best possible job he could even though learning does not occur. There are, of course, rules of strategy which, if followed, tend to produce the desired learning. Unfortunately these rules do not guarantee success if followed.

The situation is analogous to mathematics where there are rules which enable us to evaluate a purported proof. Once it is exhibited, we can determine in a precise way by following the rules whether or not the proof is a valid one. On the other hand there do not exist any rules such that if they are followed, the successful production of valid proofs is guaranteed. Furthermore this can be demonstrated conclusively on mathematical grounds alone. The case is similar in all scientific inquiries where theories once proposed can be tested, but the production of such theories cannot be guaranteed by following some set of rules. The rules for scientific inquiry are at best heuristic. Of course, these heuristic rules can be made to guarantee success by simply adding such things as "Find an appropriate theory" or "Set down a valid proof," but the addition of such "rules" is obviously no help since it is precisely because we have no idea other than that given us by the heuristic rules how to follow these "success-guaranteeing" rules.

There are, of course, rules for some activites which not only guarantee success, but are also helpful. Spelling rules are of this type. A rule to spell 'cat' might go "Leaving a letter wide space to the left write the letter 'C', then leaving no letter wide space write the letter 'A', then. . ." This rule is helpful because a

student who didn't know how to spell 'cat' might very well be able to follow the rule and if he did, success would be guaranteed.

Unfortunately teaching involves many more activities like scientific inquiry than it does activities like spelling. Thus we can independently of success in learning evaluate teaching as an activity in cases where the teacher is teaching things like spelling. If the teacher gives the rules, then he has done all he could. But even this is too strong; for the teacher must also have attempted to make sure that the students understood the rules and such understanding is <u>not</u> guaranteed by following another set of rules. The analogy is then this: Rules for effective teaching are like the heuristic rules for producing fruitful scientific theories or valid proofs. They are helpful, but do not guarantee success. Successful learning of what has been taught is analogous to evaluating a theory or checking a proof. Both show that the activity has been successful, but neither tells us how to make the activity successful in the future. Scheffler uses this analogy to discredit attempts to characterize and teach the methods of teaching in terms of rules to be followed or distinctive actions to be performed by the teacher. The most that can be done in teaching people how to teach is to provide them with heuristic rules, and, as I shall now argue, a

firm grounding in what it is they are supposed to teach.

Scheffler argues that not only is teaching a practical activity carried on with certain intentions, but also an activity carried on in a certain manner or way. Not all activities carried on with the intention of getting people to behave in certain ways can be classed as teaching. We naturally want to exclude the use of bribery, force, drugs, hypnosis, etc. from the concept of teaching. Scheffler, however, narrows the manner in which the activity of teaching is carried on even further. He argues that the manner appropriate to teaching is the free, frank, and open discussion of the reasons for and against believing anything. The teacher must teach in a way that acknowledges the reason of the pupil. The teacher must be willing and able to give his reasons to the pupil, to explain a point, to demonstrate a skill, to show the purpose in learning certain kinds of things.

Now although this is certainly a laudable goal for teachers to strive to attain, it seems to provide a rather "high" definition of teaching. There seems to be no logical reason for demanding that a person "acknowledge the reason of the pupil" in order to be called a teacher; for all that there are practical and moral reasons for him to do so. Nevertheless Scheffler's connection of manner with the activity of teaching is extremely important even if we do

not demand that the manner be quite so praise-worthy as Scheffler makes out. The importance of the manner of carrying on the teaching activity can perhaps best be brought out by considering the three learning locutions.

If the teacher is merely trying to get the student to learn to do certain things in the sense of 'learning to' which involves a high probability of the student's doing what he has been taught whenever he is in certain situations, then there seems to be much less restriction on manner than in other cases. In these cases where the goal is merely a 'learning to', the teacher tends to proceed with a minimum of explanation and a maximum of drill. He might motivate the student with threats of reward and punishment and not by trying to appeal to any intrinsic motivation the student might find in what is to be learned. He is trying to get the student merely to act in a certain way in certain situations. And, of course, there are many occasions where such a manner of teaching is perfectly acceptable. The teaching of minimal forms of courtesy, memorizing the multiplication tables, holding one's breath while under water, etc., all tend to be taught, and rightly so, in this manner. We are not concerned in these cases that the student behave intelligently--that will come later if it is needed at all--rather we are merely concerned that the student does a number of things without thinking so we

can get on to more important business.

'Learning that' is also connected with considerations concerning the manner in which teaching is conducted. If we are trying to teach a student that such and such is the case, we are, of course, trying to get him to believe it, but we may only want to teach him to say that such and such is the case in certain situations. This sort of 'teaching that' is thus assimilated to the 'learning to' discussed above, and may well include the methods discussed there. On the other hand we may teach the student that such and such is the case for reasons, but entirely the wrong kind of reasons. The student may learn that Columbus discovered America because he has learned that the teacher will very probably ask for this information on a test and will mark the student down if he cannot provide the information. Finally there is the 'teaching that' which is directly connected with the student's reasons in 'learning with understanding that'. In this case the teacher's manner must include the casting of points in terms the students can understand so that the students' reasons are not only the teacher's reasons for the proposition, but also the reasons for believing the proposition. This last is the manner which Scheffler is referring to in saying that the teacher's manner in his teaching activity is the acknowledging of the student's "reason." To teach in this manner will

probably involve discussion and question-answering, but the precise method is not important. The important thing is that to teach in this manner is to be prepared to submit the teacher's reasons for a belief to the student's evaluation and criticism. It is clear that if the teacher intends his students to learn with understanding, he must be prepared to submit his (the teacher's) reasons for the student's judgment.

'Learning how' is, as I have already argued, composed of both 'learning to' and 'learning that'. As such, the proper proportions of these latter two must be kept in mind when considering the proper manner of 'teaching how'. To teach how is, in general, to show how and explain how in addition to teaching that such and such are the relevant considerations for the application of the skill being taught. Thus the manner of 'teaching how' may be compounded of the manner of 'teaching to' and the manner of 'teaching that' in varying proportions depending both on the teacher's intentions and the particular skill to be taught.

The problems involved in not keeping the intentions of the teacher as to what kind of learning is to be fostered consistent with the manner of teaching ought by now to be apparent. To take an extreme case, suppose a teacher intends to produce learning with understanding (as most would claim they do), and then proceeds to teach in a manner

appropriate to 'learning to'. He gives the students excessive drill, puts a high premium on their memorizing his exact words, discourages questions from the students, and in general behaves as if he were conditioning animals and not teaching persons. It is not, as Scheffler would have us say, that the teacher is not teaching; it is rather that he is teaching his students to do and say certain things when he ought to be teaching them that such and such is the case in the sense of their learning this with understanding.

There are also examples emphasizing the undesirability of too much of the 'teaching that' manner. A teacher who spends too much time meeting all the students' objections, questions, etc. on matters which ought only to be the subject of 'learning to' is at best wasting everyone's time and at worst giving his students a distorted sense of the relative importance of what is to be learned. The problems involved in 'teaching how' are also clear. If the teacher is not aware of the proper "mix" of 'learning to' and 'learning that' involved in the acquisition of the particular skill he is trying to teach, he may emphasize the wrong things to the detriment of all concerned. Thus he may spend too much time on routine practice to the exclsuion of some of the 'learning that' which must occur if the skill is to be more than a rote piece of behavior.

Or he may spend so much time explaining that the students never do have a chance to practice the skill at all.

Of course, the most efficient relation between the intended kind of learning and the manner of teaching is to a very large extent an empirical matter, and I do not presume to be doing a priori science in the foregoing. Rather what I have been arguing is that the students very often read off a teacher's intentions as to the learning he is attempting to foster from the manner in which he teaches. The teacher who professes to be interested in the students' learning with understanding but who teaches and tests his students in a manner to encourage mere 'learning to' should not be surprised that he often fails in his professed goal. This is why the teacher who wants his students to learn with understanding must be a master of what he teaches. For if, as I have argued, learning with understanding is primarily a matter of coming to have good reasons for a belief in the way my analyses have indicated, then the teacher must know his subject so well that he can give the student the reasons which are reasons for the belief being taught and learned. The teacher who has good intentions as to the type of learning he wishes to foster but who does not have a mastery of the subject is at a serious disadvantage. Too often is he able only to foster 'learning to' and not 'learning with understanding that' because he does not

himself understand the reasons for what he is teaching and thus a fortiori cannot lead the student to any sort of understanding. He teaches the student to do and say certain things in certain situations, not that such and such is the case because of such and such a reason.

The foregoing remarks, if correct, indicate a certain mutuality of endeavor between student and teacher if true learning with understanding is to occur. The teacher must be prepared not only to set forth his reasons, but to have them evaluated and criticized by the student, and possibly the teacher may even have to admit that he is wrong. The student on the other hand must not only learn to repeat at the appropriate times what the teacher has said, but must question and study what the teacher says and does and what his reasons are for saying what he says so that he, the student, may come to have his own reasons for or against any given belief. This mutuality of interests is especially important in moral education where the goal is to get the student to behave in certain ways in the way he has learned how to behave because he has learned that one ought to behave in these ways. The student who is not allowed to question conventional morality and is not allowed to question the justification for conventional morality will never learn to be a moral person in the full sense specified above; for he will never learn with understanding

that one ought to accept certain moral norms or principles. If he does learn to be moral in spite of his instruction, it will be much, much harder.

Thus Scheffler is wrong in saying that fostering learning in a manner that acknowledges the student's reason is a logically necessary condition of an activity's being describable as teaching, but he is right in saying that the activity of teaching is necessarily carried on in certain fairly determinate ways and these ways are at least empirically connected to the kind of learning which is likely to be fostered. Scheffler's "manner" is an ideal manner--one devoutly to be wished for in many cases--and the one most likely to encourage learning with understanding. It is not, however, an easy method to adopt; for it requires a high degree of mastery of the subject matter by the teacher and a willingness to be proved wrong occasionally, both of which attributes too few teachers possess. Finally the question of the manner of teaching has been seen to be closely related to the kind of learning intended; and if nothing else, attention to a clear formulation of the kinds of learning to be sought and their relation to the manner of teaching ought to improve immeasurably our understanding of the educational process and how to improve it in addition to pointing out promising avenues of empirical research to establish just what the best manner of teaching for different kinds of learning is.

CHAPTER IV

THE PSYCHOLOGICAL "REDUCTION" OF LEARNING LOCUTIONS

It is quite obvious that the analyses I have given of the three paradigm learning locutions--'learning that', 'learning how', and 'learning to'--are in terms of primitive concepts which have variously been characterized as "mental," "intentional," "psychological," etc. For example, both 'learning that' and 'learning how' explicitly make use of the concept of "belief" which is often taken as the paradigm case of a mentalistic concept. 'Learning to' as well as the others involves "intending" in the exercise of what has been learned, for it is this intending to perform certain actions which distinguishes rote learning from mere mechanical conditioning. Now there are various psychologists and philosophers who would claim that such analyses in terms of mentalistic concepts are worse than useless. They are positively false and misleading. These people would claim that statements in terms of mentalistic concepts are to be eschewed in favor of more scientific language about what can be empirically verified, for in principle mentalistic concepts are somehow "ghostly" or at the very least "private" and thus not

subject to empirical observation. It will be the purpose of the present chapter to analyze such a "reductionistic" program and to develop some minimal linguistic criteria which any such claimed "reduction" must meet if it is to be a success.

1. A Note on Method

I must first make a rather rough and ready distinction in the philosophy of language without delving too far into this subject and without settling conclusively any of the controversies which may still surround the distinctions I am going to make. The first distinction I wish to make is that between the meaning and the use of a sentence. Very roughly, the meaning of a sentence has to do with the thought that is expressed, the cognitive content of the sentence. The meaning of any sentence is somehow constituted by the meanings of its parts, and these in turn are more or less definitely fixed for the speakers of any natural language. Thus there is a real limitation on what a speaker of language can mean by his words. This limitation is that he cannot simply intend his words to mean something other than they do mean and hope to be understood. Thus a man cannot say "Grass is red" and hope to be understood as uttering a true sentence simply because he intends by his use of the word 'red' to mean what ordinary speakers of English mean by 'green'. He would at least have to

make clear to his audience that he was using 'red' with the meaning that 'green' is ordinarily taken to have, and this "making clear to his audience" would have to be with words which themselves have their ordinary meanings or else his audience would have no idea what it was he was explaining to them. Briefly the words, expressions, and sentences of a language must have more or less precisely fixed meanings for all speakers of the language in order for the communicative purpose of the language even to get off the ground. And this remark would have to be true no matter how 'meaning' is eventually to be analyzed.

Thus, I take it that a theory of "meaning" would have to include a theory something similar to the Fregean theory of sense and reference. It would include a specification of how the sense of a word or sentence determines its reference without worrying about the actual references of the words in any specific sentence, and thus without worrying about the actual truth or falsity of the sentence. In other words, knowing the meaning of a sentence involves knowing at least in a vague way the truth conditions for the sentence. An adequate theory of meaning thus might include the theory of representing natural language sentences in modern quantification theory to exhibit their logical structure, the theory of sense and reference, Tarski's theory of truth, and perhaps a theory of opaque

contexts as well. In short, "meaning" is taken to be that which must be minimally understood if one is to "understand" a word or sentence in a natural language, regardless of the actual truth or falsity of the sentence and regardless of the intentions of any speaker in using any particular sentence. Theory of "meaning" as I am using 'meaning' would concern itself with sentence "types" and not sentence "tokens." Roughly, the distinction between a linguistic type and a token of that type is that the former is a linguistic entity and the latter is a physical object of some sort. Thus in the following box, there is but one sentence type, but two sentence tokens.

Grass is green
Grass is green

However, once one moves from a consideration of the linguistic properties of sentence types to a consideration of sentence tokens as well, then one has entered the realm of what I want to call "use." Sentence tokens occur in specific contexts and are uttered or written with particular intentions in mind. Often these contexts are such that we are interested in the truth or falsity of the particular sentence token. In such cases, from our knowledge of the "meaning" of the sentence, i.e. our knowledge of the truth conditions of the sentence, and our knowledge of

the specific context of use, we can determine the actual truth or falsity of the sentence.

Although the assertive use of sentences is indeed a common one, i.e. the use of sentences to make a truth claim, it is by no means the only use of sentence tokens. That is, although the words and sentences of a language have a more or less fixed meaning, there are any number of uses to which these words and sentences may be put. Thus I may utter the sentence, "The door is shut," intending to get my audience to believe that the door is shut. This is a more or less informative or assertive use. However, I may have the further intention of warning my audience that they must not simply try to back out of the room. This further intention could probably have been fulfilled if I had not used language at all. I might have simply grabbed my "audience" by the arm and turned him around so that he could see that the door was shut and would not blunder into it. On the other hand I may utter the sentence, "The door is shut," intending to ask a question. Ordinarily I would use a particular tone of voice to do this. I may also have had the further intention of getting my audience to open the door for me if it is shut. Again I may utter the sentence, "The door is shut," to my son as he comes bursting into the house leaving the door open intending by my utterance that he

understands I wish him to shut the door. This utterance is also usually made in a particular tone of voice, and I may have the further intention of reprimanding him by uttering it. Finally I may simply use the sentence, "The door is shut," as an example of an indicative sentence in some class I am teaching with no more intention than that my audience take it as an example. There are, of course, an indefinite number of further ways I might use the same sentence, and examples can be easily constructed of these further uses.

However, not only are there an indefinite number of different uses of the kind I have just been considering, there are also an indefinite number of hierarchical uses as well. For example, one could consider all the above cases being spoken by a character in a play or novel. The uses outlined above probably are still operative in <u>some</u> sense or other, but there is added on to these the special use of the words as being used in the context of the dramatic or literary work. Or consider the case of lying or intentionally misleading someone. Again the situations might be the same and perhaps even most of the intentions. But here too there is <u>another</u> use to which the sentence is being put. That is, the liar is using the sentence with the intention of getting someone to believe something which is not the case. Or consider the metaphorical,

allegorical, and simile uses of sentences. Surely there is no special "meaning" involved here for we understand perfectly well what the words mean and it is only because we understand their meaning that we realize that the use is a metaphorical, allegorical or simile one. In short, I take the "use" of a sentence to refer to the particular context of its utterance and the particular intentions which the speaker has in uttering it.

However, I must make a further distinction between two different kinds of uses or intentions--the public and the private. Many of the uses and intentions that I have illustrated above are what might be called "private" intentions. What this roughly involves is that I had intentions in uttering what I did which could have been fulfilled even without my using language at all. Thus I had respectively the intention of warning my audience that the door was closed, of getting the audience to open the door for me, of reprimanding my son, etc. There were also, however, certain "public" intentions or uses of language which also entered into the linguistic act I performed. This "conventional" use of language is probably best illustrated by two examples.

The most common "conventional" use of language is probably the assertive one. If I say in normal circumstances and a normal tone of voice, "The door is shut," and I make

use of no special conventions that indicate I am joking, using the sentence as an example, etc., then I can be said to have asserted that the door is shut. Furthermore, if the door turns out not to be shut, I am held prima facie responsible for having misled my audience with my assertion. And it is no defense for me to claim that I did not intend to assert that the door is shut if in fact I made use of the normal conventions for asserting without explicitly qualifying them in any way. I am held responsible for having asserted, and I gave my audience the right to assume that I was asserting. I thus lay myself open to a certain sort of censure if I did not actually intend to assert.

Another example is that of promising. As a matter of fact whenever I say "I promise. . ." in suitable circumstances, I have promised. And this is so despite any intentions I may have had about not wanting to promise or not keeping the promise. My audience is perfectly justified in holding me to my promise if I have not indicated in some way that I was joking, or merely expressing an intention to try to perform, or pretending to promise, or one of a number of other things which effectively remove the conventional force of "I promise. . . ." This is so because the usual private intention in uttering the words, "I promise. . .", is that of placing oneself under an obligation. Because of the importance of this private

intention, it has become institutionalized and thus a "public" intention for which people are entitled to hold me responsible despite my varying private intentions. That is, the fact that whenever someone utters the words, "I promise. . .", in suitable circumstances, he has promised is a public conventional use of these words.

It will be observed that my distinction between meaning, public use, and private use corresponds very closely with John Austin's distinction between locutionary, illocutionary, and perlocutionary descriptions of a speech act.[1] However, Austin seems to believe that the fact that whenever an English-speaking person says, "I promise. . ." in appropriate circumstances, he has promised is a part of the meaning of "I promise. . . ." This will not do. What these words mean is that I am or am trying to obligate myself to perform what I am promising by going through the appropriate ritual or convention for doing this, whatever that ritual might be. As a matter of fact, one way this ritual may be gone through with English-speaking cultures is by uttering these very words. But this need not be so. We could very easily imagine a culture where the ritual of promising consists merely in the crossing of the fingers of the left hand whenever someone

---

[1] J. L. Austin, How To Do Things With Words, (Cambridge, Mass.: Harvard University Press, 1962).

asks if you will do something. The promiser need not speak any words at all. And if he did, he might merely be saying that he was trying to go through the appropriate ritual. Thus the words "I promise. . ." can be used in English to promise independently of their meaning and independently of any other private intentions. Moreover, this use is a "public" one.

In summary then, for me, "meaning" is concerned with what must essentially be known of a language in order to understand that language. "Public intentions" have to do with the conventions of a language which as a matter of fact are associated with the language, but which we could imagine being different in form even if we could not imagine that there are not some such conventions. The "private intentions" refer to those further intentions which language users may have, but which could in general be fulfilled without using language at all. I admit that this distinction is only a rough and ready one, but I believe that it will do for my purposes.

However, I must try to meet one possible objection which seems to be currently popular. This objection, often voiced by philosophers of a Wittgensteinian persuasion is that meaning is use. That is, if we know all there is to know about the use of language or the intentions with which we use language, then we know all there is to know about

meaning. In order to meet this objection I must first make two concessions. First, I grant that the meanings of words do depend on their use. But this dependence is on the history of their use and does not at all imply that the present meanings of words could be analyzed in terms of their use on any particular occasion. It is indeed a contingent fact that 'red' means red and not green. It is possible that 'red' and 'green' might have had exactly opposite meanings. But this fact goes nowhere toward showing that this is how we must analyze the present day meanings of words. In fact, I shall argue that the reason we can use words and sentences in the way we do is precisely because they have the meanings they do.

The second concession I am willing to make concerns my notion of "public intentions." It might be possible to subsume public intentions either under meaning or under private intentions, although no such attempt has yet succeeded. I have already argued in the case of promising that public intentions cannot be analyzed in terms of meaning. The argument for its not being analyzable in terms of private intentions has also been given in distinguishing public from private intentions. In essence this argument depends on the fact that when conventions (public intentions) are used without qualification, the audience is entitled to hold the speaker "responsible" in some sense for his

utterance <u>despite</u> any private intentions he might have. However, I am willing to grant that a successful subsumption might eventually be effected for "public intentions." At the present time, however, I believe the distinction I have made is still a valuable one.

But what about "use" in general providing necessary and sufficient conditions for an analysis of meaning? The following argument is not meant to be absolutely conclusive since it depends rather heavily on some controversial aspects of the philosophy of language, and any attempt to make it conclusive would involve me in setting out a fairly complete theory of language. This undertaking would require a book in itself. However, the argument and the conclusions I will reach will enable me to make some enlightening comments on the problem of the "reduction" of ordinary learning notions to psychological learning theory.

The proponents of the "meaning is use" doctrine would begin by saying that all my talk about meanings and intentions is totally mysterious. What must be done, they argue, is to use the public, observable knowledge that we have and simply look at all the situations in which people use language. Once we have done this, then we can simply define meaning in terms of the content of use by people. We thus do away with mysterious entities in favor of observables.

Such a position is very hard either to attack or to defend, for it is a well-nigh impossible task to observe with any detail all the contexts of language use, and even if we could, it is not clear how this would help. The proponents of this view, however, make the following claim to bolster their case. They say that we can study a few simple "language games" (the term is Wittbenstein's), and notice all the ways in which the words are being used and this will constitute knowing the meaning of them. Then by analogy, so the argument runs, we will understand the complex language game being played by the users of a natural language. Now by "language game," the proponents of this view seem to mean simply a complete description of the activities and context in which language is used.

Consider, for example, the following language game described by Wittgenstein:

> The language is meant to serve for communication between a builder A and an assistant B. A is building with buildingstones: there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a language consisting of the words "block", "pillar", "slab", "beam". A calls them out;--B brings the stone which he has learnt to bring at such-and-such a call.--Conceive this as a complete primitive language.[2]

Now the proponents of the view under consideration seem to feel that from such a description we know all there is to

---

[2]L. Wittgenstein, Philosophical Investigations, trans. by G.E.M. Anscombe (New York: MacMillan and Co., 1953), §2, p. 3e.

know concerning the meanings of the words involved in the language game. We know their complete use and what more is there? However, I believe that such a description of the use of the words of this primitive language is neither a necessary nor a sufficient condition for an analysis of 'meaning'.

To show this, suppose the builder yells, "Slab!", and the helper decides to take a coffee break. On Wittgenstein's view we cannot say that the helper understands, because his behavior does not fit into the game. But surely he does understand, thus showing that this particular kind of behavior is not a necessary condition of a word's having meaning. On the other hand, what if the helper believes that those sounds coming from the builder are natural expressions on the part of the builder when he has an overwhelming desire for that type of stone without any intent on the builder's part to communicate this desire to the helper. Then the helper, being a good-hearted person, brings the appropriate stone. In this case although all the conditions are satisfied, it cannot be said, I think, that the helper knows what the builder's utterances <u>mean</u>, if they mean anything at all.

The point I have been stressing in arguing the non-collapsibility of meaning and use is that an important class of intentions a speaker has are the intentions to

fulfill his further private intentions _through_ the audience's understanding the meanings of the words which the speaker utters. Thus the builder intends his helper to bring a slab whenever the builder says "Slab!", but further, he intends that the helper bring the slab through a recognition of the meaning of "Slab!", and a knowledge of the conventions governing commands or requests. Thus meaning is not to be identified with use, at least not in the behavioristic way in which Wittgenstein would have us believe.

Before leaving this topic, I want to reemphasize two points. First, I take my arguments separating use and meaning to be conclusive only against certain proposed identifications of these two concepts such as the one connected with Wittgenstein's simple language games. These arguments also _seem_ to provide a clue as to where to look for a possible mistake in any future proposal of this type. I am _not_ claiming, however, that no _possible_ identification of meaning and use will ever succeed, only that none so far have. Second, despite the apparent independence of meaning and use, there is still a very close connection between them; and, at the very least, a word with a certain meaning has a more or less specifiable _range_ of uses to which it may be put, and conversely, a certain range of uses indicates a certain family of meanings. If

this is so, then any claim of synonymy between the two linguistic expressions entails a roughly similar range of uses and a widely dissimilar range of uses for two linguistic expressions entails at least a serious doubt as to their synonymy. These points will become extremely important in the following sections.

2. The Development of Some Linguistic Criteria for a Successful Reduction

The concept of "reduction" is most clearly applied in the field of science. A reduction of one theory to another occurs roughly whenever the theory being reduced (secondary theory) is explained in terms of the other theory (primary theory). This usually involves the deduction of the secondary theory from the primary theory. In general, the two theories are concerned with prima facie different subject matters; for if they are not, the "reduction" simply seems to be an "expansion" of a given theory and usually causes no philosophical problems. For my purposes, "reduction" will refer to any attempt to replace a particular class of statements in ordinary language by another class of statements in ordinary language or in a scientific theory.

Before considering a few examples of a proposed reduction and some arguments against it, it will be helpful to state a criterion for a successful scientific

reduction which seems to apply here.[3] The criterion is that certain assumptions must be made connecting the terms of the secondary theory which do not already appear in the theoretical terms of the primary theory with certain of the theoretical terms of the primary theory. That such assumptions are necessary is clear. For by principles of logic we cannot (except in special cases which are not applicable here) derive conclusions containing terms which were not contained in the premises. Thus if the secondary theory contains a term, 'A', which is not present in the theoretical terms of the primary theory, we could not derive or explain the statements of the secondary theory containing 'A' using the statements of the primary theory unless such linking assumptions are made. Now these linking assumptions seem to be of three and only three kinds. First they might be logical or meaning relations between the terms of the two theories (in which case this is not really an "assumption"). Secondly, the linkages may be in the form of

---

[3]Ernest Nagel, The Structure of Science (New York: Harcourt, Brace and World, Inc., 1961), Chapter 11. In a strict sense I am not talking about a "scientific" reduction here. A truly "scientific" reduction can occur, according to Nagel, only when the axioms, experimental laws, special hypotheses, etc., of both the primary and secondary theories are more or less explicitly formulated. Otherwise, it is very difficult to decide with any assurance whether the proposed reduction is successful or not due to the vagueness and ambiguities involved. Such precision is clearly lacking in the cases with which I shall be concerned. Nevertheless, some of the criteria for a successful scientific reduction can be applied, it seems, to the sort of reduction of which I shall be speaking.

empirical laws which as a matter of fact link the occurrence of a state of affairs described in the secondary theory to another state of affairs described in the primary theory. In this case the law must be empirically supported and the two states of affairs must, in principle, be capable of being independently identified. Finally, the linkages may be stipulative only, and thus created by deliberate convention. This case can arise if the term (or construction of terms) in the primary theory has no empirical significance other than that created by the stipulation or has the same empirical significance as the term in the secondary theory to which it is linked.

A notorious example of an attempted reduction of the type I want to consider is the thesis of "naturalism" in ethics and value theory. This thesis is that there are two realms of discourse in ordinary language, the factual and the evaluative, which prima facie seem to be talking about two different sorts of things. However, the thesis goes on to claim that sentences involving evaluative terms can be replaced by sentences involving merely factual terms. In short the holders of a naturalist position in ethics claim that evaluative notions can be reduced to factual ones.

The kind of linking assumption which is usually taken by naturalists as the appropriate way to link evaluative and factual terms is the logical or meaning one.

That is, most naturalists claim that statements containing evaluative terms mean the same as other statements containing only factual terms. For simplicity I shall consider only the case of reducing sentences containing 'good' to sentences containing merely factual terms. Thus perhaps the most famous naturalistic position, utilitarianism, might be construed as claiming something like 'X is good' means 'X is productive of more pleasure than pain' where 'pleasure' and 'pain' are understood to be factual terms, i.e., the states of affairs described by these terms can be identified by purely scientific means. The exact details of such an identification are immaterial for my purposes, but we might suppose that things are pleasurable just when certain circuits in the brain are triggered (a supposition not at all inconsistent with current brain physiology). At any rate the question of whether something is good or not is supposed to be capable of being settled in principle by an appropriate scientific investigation.

Of course, utilitarianism is not the only possible naturalistic definition of evaluative terms. The following also qualifies: 'X is good' means 'A majority of people believe X is good'. Ignoring for the moment the problem of ascertaining scientifically peoples' beliefs, this too can be construed as naturalistic. In principle we might find out if something is good or not simply by asking

everyone whether they believe it is good. If a majority of them do, then it is good. Notice in this case that the predicates 'is good' and 'is believed to be good' are both one-place and the former is dependent for its meaning through the asserted equivalence on the latter. The latter might be primitive or might be analyzed further. If it is further analyzed, it is implicit that the analysis will not contain 'is good', but only further factual terms.

Although at one time or another and in one form or another such "reductions" of evaluative notions to factual ones have been very popular and persuasive, they have also come under serious attack. Recently this attack has centered around Moore's famous "open question" refutation of naturalism and various reformulations of this argument. It will be instructive to examine in some detail one of the most recent and plausible formulations of this argument. Such an examination will enable me to extract some general criteria for a successful reduction of this kind in terms of the meaning-use distinction already discussed.

R. M. Hare in his influential book, The Language of Morals, formulates the argument as follows:

> Let us suppose for the sake of argument that there are some 'defining characteristics' of a good picture. It does not matter what sort they are; they can be a single characteristic, or a conjunction of characteristics, or a disjunction of alternative charactristics. Let us call this group of these characteristics C. 'P is a good picture' will then mean the same as 'P is a picture and P is C'. For

> example, let C mean "Having a tendency to arouse in people who are at that time members of the Royal Academy (or any other definitely specified group of people), a definitely recognizable feeling called "admiration"'. The words 'definitely specified' and 'definitely recognizable' have to be inserted, for otherwise we might find that words in the definiens were being used evaluatively, and this would make the definition no longer 'naturalistic'. Now suppose that we wish to say that the members of the Royal Academy have good taste in picture. To have good taste in pictures means to have this definitely recognizable feeling of admiration for those pictures, and only those pictures, which are good pictures. If therefore we wish to say that the members of the Royal Academy have good taste in pictures, we have, according to the definition, to say something which means the same as saying that they have this feeling of admiration for pictures which have a tendency to arouse in them this feeling.[4]

One paragraph later Hare generalizes this argument as follows:

> If 'P is a good picture' is held to mean the same as 'P is a picture and P is C', then it will become impossible to commend pictures for being C; it will be possible only to say that they are C. It is important to realize that this difficulty has nothing to do with the particular example that I have chosen. It is not because we have chosen the wrong defining characteristics; it is because, whatever defining characteristics we choose, this objection arises, that we can no longer commend an object for possessing these characteristics.[5]

However, this argument, plausible as it may sound, rests on a series of related assumptions which are not specified by Hare, and which are dubious at best. Let me examine these. First of all it is clear that "commending"

---

[4]R. M. Hare, The Language of Morals (London: Oxford University Press, 1961), p. 84.

[5]Ibid., p. 85.

something for possessing characteristics, C, is a use of 'good' in the sense of the meaning-use distinction formulated above. For my present purpose it does not matter whether the commendatory use of evaluative notions falls under the public intentions category or the private intentions category. My own feeling is that it falls under the former, but this is not important. What is important is that Hare seems to assume that the meaning of an expression can be identified with its use; for all his argument has shown is that if this form of naturalism is true, we seem to be barred from using evaluative expressions in certain ways. Unless use is meaning, this does not show that a naturalistic thesis is false; for naturalism connects meanings, not uses.

Although I argued against an identification of meaning and use in the foregoing section, I admitted that it might just be possible. Let me suppose now that such an identification is possible. What is the status of Hare's argument in that case? Well, it seems that if meaning is use, Hare's argument is almost conclusive. I say "almost" because it still seems to be the case that there may be some uses of an expression which are not central to the meaning of the expression. This seems to be the case with most uses falling under the private intentions category, although not with those falling under the public

intentions category. The reason for this is that private intentions can in general be satisfied without using language at all. Thus it is hard to see how they could contribute to the meaning of a linguistic expression. On the other hand public intentions seem much more closely bound up with languages and it is probably here that any identification of meaning and use would have to take place. Thus Hare needs to supplement his argument by showing that the commendatory use of evaluative expressions is really quite central to their meanings. I do not think that this would be difficult, and it would probably be admitted by naturalists, but it seems Hare would have to do this to make his argument complete.

But what if meaning and use cannot be identified? If this is the case, then it seems to me that Hare's argument shows at most that there is a different range of uses for evaluative expressions and factual expressions. Since, as I admitted, there is even in this case a very close connection between use and meaning, Hare's argument would thus cast a serious amount of doubt on the naturalist thesis depending on just how central the commendatory use is. The naturalist might again save his claim by arguing that the commendatory use of evaluative expressions is not one of the uses closely connected to the meanings of evaluative expressions. However, if, as appears likely,

the commendatory use is central and would thus have to be taken account of by any naturalistic definition of evaluative expressions, Hare's argument may be conclusive against that particular form of naturalism which he considers. It does not show, as Hare claims it does, that no naturalistic definition will do. There seems to be no a priori reason why some factual terms may not have commendatory uses in certain contexts. It thus seems that it is still possible for some naturalistic thesis to be correct. The only requirement is that the factual terms of the proposed definition be capable of being used to commend, blame, praise, advise, and so on. Hare's argument is typical of anti-reductionist arguments. A case can often be made out that the arguments are conclusive for particular instances of the proposed reduction, but when an attempt is made to generalize these arguments, the generalization is often not valid. In other words, the situation is not analogous to that of the foundations of mathematics. In any system strong enough to contain elementary number theory, Gödel has proved that there can be no reduction of the concept of mathematical truth to the concept of provability in the system. Such a situation does not in general obtain for other controversial reductions, i.e there is no general proof that the reduction must fail.

The foregoing discussion seems to provide the basis

for formulating some linguistic criteria of a successful "reduction" of one class of statements to another if the linking assumptions are taken to be true in virtue of the meanings of the words alone. Whether meaning is identical with use or not, it seems to be the case that any sentence of the secondary theory, must have roughly the same uses as a sentence in the primary theory which is logically equivalent to it when the appropriate definitional replacements have been made; and conversely.

If meaning <u>is</u> identical with use, then pointing to a use of a sentence in the secondary theory not possessed by the logically equivalent sentence in the primary theory casts serious doubt on the reduction. However, such an anti-reductionist argument has to be supplemented by showing that the use is "central." This may be admitted by the proponents of the reduction, but if it is not, then an argument showing that a significant part of discourse about the subject matter is unaccounted for must be produced in order to show the reduction to be fallacious.

If meaning is <u>not</u> identical with use, then pointing to a use in the secondary theory not possible in the primary theory again casts serious doubt on the proposed reduction. However, in this case the proponent of the reduction can escape the argument not only by showing the use not to be central, but also by admitting that the

particular proposal is indeed erroneous. He may nevertheless claim that the anti-reductionist argument is not general but merely telling against the particular form of the reduction he has proposed. Furthermore, the argument can then be used as a guide to finding a new linking definition which does have the use in question. Since this sort of move seems usually to be open to the reductionist, it appears that conclusive general anti-reductionist arguments are not often possible. Each proposed reduction must be judged on its own merits and even a series of failures may not give any support to the position that no reduction of the kind in question is possible.

I now want to turn to a consideration of the second kind of linking assumption-empirical laws. Let me suppose, as very few naturalists would, that the linking assumptions between evaluative discourse and factual discourse are really only empirically true. For example, it might be claimed that all and only those things which are "good" are also "pleasant," although it is admitted that synonymy is not in question here. It is evident that in this case independent evidence must in principle be obtainable for the state of affairs signified by 'good' and for the state of affairs signified by 'pleasant'. If this were not so, then it would be impossible to either confirm or disconfirm this hypothesis. It is probably for this reason that no

naturalist makes a claim of this kind, for it is very hard to see just how the evidence for the application of 'good' would differ from evidence for the application of 'pleasant'. However, this is immaterial to my present purposes.

If this is the kind of link postulated, then it seems that the meaning-use distinction is no longer applicable, for it is granted that the terms have different meanings and therefore a different range of uses. Indeed it does seem to be the case that the only help a consideration of meaning and use could have here would be to clarify the terms involved to make it plain that they really do have independent criteria for their evaluation. Such considerations might save psychologists the pains of testing analytic statements, but that would be all.

However, this in itself would be a great deal. For example, a widely used test in educational psychology makes the following statement.

> The pupil who learns a prescription without understanding it will have difficulty in adapting it to a new situation. He may learn rapidly a procedure presented in a rote fashion. But the Brownell-Moser study shows that he is then less able to work out the proper procedure in a slightly different situation, and less equipped to understand subsequent explanations. Teaching materials that can be understood and teaching methods that foster understanding are needed if learning is to transfer.[6]

---

[6]Lee J. Cronbach, Educational Psychology, 2d. ed. (New York: Harcourt, Brace and World, Inc., 1963), p. 347.

And then in the next chapter entitled, "Improving Understanding and Thinking," the author says,

> The only sure sign that the pupil has command of a principle is transfer; he operates with the principle, going beyond the words given by text or teacher and beyond the tasks he has previously practiced.[7]

A reasonable interpretation of this would be to say that understanding means the same as transfer, but if it does, it is hardly surprising that "Teaching materials that can be understood and teaching methods that foster understanding are needed if learning is to transfer," for all this says is that what promotes transfer promotes transfer. It is only if "understanding" and "transfer" can in principle be independently found to obtain that the above is an empirical hypothesis and this is just what Cronbach does not make clear.

Nevertheless, considerations of meaning and use do arise in another place even if the postulated link between the terms of the two theories is actually a legitimate, true, empirical hypothesis. This sort of situation may arise some time after such a reduction has been successfully carried out. It is sometimes argued that words do change their meanings, albeit slowly. Thus, although 'A' may at one time have meant something different than 'B', because of the fact that 'All and only A's are B's' is an

---

[7] Ibid., p. 359.

empirical truth and was used to effect the reduction, it is sometimes said that 'A' now means the same as 'B'. It is here that meaning-use considerations become relevant to assessing this claim. If the original meaning and uses of 'A' have been entirely or almost entirely discarded, then it might be properly claimed that 'A' is not even ambiguous, i.e. it no longer even means what it used to as well as B--its present meaning. This, I suggest, is the explanation of what has happened to the term 'atom' in the history of science. When the early atomic theory was reduced to the present one, there may have been a time when 'atom' was empirically correlated with expressions using only the terminology of the primary theory. This may have come later to be considered a <u>definition</u> of 'atom', and today it is fairly clear that 'atom' does not mean an indivisible particle, but a certain structure which minimally determines the properties of a chemical element. (Of course other meanings of 'atom' or 'atomic' can be stipulated, e.g. an atomic sentence in the propositional calculus.)

It might on the other hand be the case that 'A' is still used to mean what it did in terms of the uses it had in the secondary theory and yet sometimes be used as meaning the same as 'B'. In this case some sort of decision must be reached as to whether 'A' means the same as 'B' or

not, and it is by meaning-use considerations that such a decision is reached. If the uses and meaning of 'A' in the secondary theory are still important or are simpler or more economical for certain contexts, then 'A' does not mean the same as 'B', and uses which tend to call them synonymous are to be avoided. On the other hand if the meaning and uses of 'A' in the secondary theory are not essential or important, then 'A' has its original meaning and uses only by courtesy for the purposes of the history of science. The historian should say something like "Although 'A' now means 'B' it once meant 'C' and through such and such a process it gradually came to have its present meaning."

I would suggest that 'temperature' illustrates the former case. At the present it still has the meaning it had in thermometry and classical thermodynamics despite the reduction of classical thermodynamics to the kinetic theory of gases by means of an empirical correspondence between classical temperature and molecular energy. This is so because concepts of mercury thermometers are still essential to the man in the street despite being inapplicable to very high or very low temperatures. However, it is conceivable given a greatdeal more scientific sophistication for everyone that 'temperature' will lose this meaning and come to mean only molecular energy. 'Temperature'

might even in the far distant future drop out of the language altogether, as 'phlogiston' has.

The point to be made here is that the meaning and uses of a term "reduced" by an empirical law must be considered whenever it is claimed that the term now "means" something different than it did. There may be all sorts of reasons for maintaining the original meaning and uses. These may include things such as simplicity, economy, ease of expression, as well as arguments showing that the original uses are somehow "essential" for the business of living. If these are <u>good</u> reasons, then it seems we must accept the fact that we have one subject matter and two vocabularies for describing it and even that this may be "essentially" so.

Finally there is the possibility that the linking postulates between the primary and secondary theories are simply a matter of stipulation. There is, of course, no objection to such a procedure and it is often a useful one, <u>as long as</u> it is made abundantly clear that the linkage <u>is</u> a stipulation. Because of the possibility of misunderstanding, words with clearly established meanings should in general <u>not</u> be used as terms for which a definition is stipulated. Furthermore this procedure is usually helpful only when there is a clear-cut criterion for the application of the terms in the definiens. It is seldom helpful

to stipulate that the meaning of 'A' is the same as 'B' where neither 'A' nor 'B' has any clear empirical application. Moreover, if a stipulation is made, 'A' and 'B' must have the same meaning and uses. We rule that such is the case, and any evidence showing 'A' and 'B' not to have the same meaning and uses is evidence to show either that the linkage is not a stipulation or the fact that it is merely a stipulation has not been clearly understood.

For example, a scientist may stipulate that by 'material object' he means a certain fairly dense conglomeration of electrons, protons, and neutrons. This may be a useful stipulation for him resulting in a certain ease and clarity of expression in expounding certain physical theories. Furthermore, there should be no problem in understanding sentences uttered by the scientist and containing 'material object' so long as this stipulation is kept clearly in mind. However, if the scientist begins saying such things as "Material objects are not solid," or "One cannot see material objects," then difficulties may arise. If it is clear from the context of these utterances that the original stipulation is being used, then although the utterances may sound strange, there is no logical difficulty.

If, however, as is often the case, the context of the utterances is such as to lead one to believe that the

scientist is trying to make some sort of philosophical claim concerning perception, for example, then it would be entirely appropriate to tell the scientist that he has ignored his stipulation and the logical requirements of it. For it is clear that he is using his stipulated definition of 'material object' as a basis for claims concerning what we ordinarily take material objects to be, and it is obvious that his stipulated definition of 'material object' is not the same as the ordinary meaning of 'material object'. This is easy to see, because we can see material objects (ordinary meaning), whereas we cannot see electrons, protons, and neutrons.

The stipulation is a bad one precisely because it leads so easily to such confusions. The scientist would be on much firmer ground if he did not stipulate a definition of 'material object', but rather made the empirical claim that material objects (ordinary meaning) are composed of electrons, protons, and neutrons. But if he did this and recognized his claim as a merely contingent one, he would not be so apt to make such foolish statements. This would be a case of two essential vocabularies with different meanings and uses to describe the same subject matter on different levels of discourse.

3. Application of the Criteria of Reduction to Psychological Learning Theory

In this section I want to try to apply some of the

linguistic criteria developed in the preceding section to certain psychological learning theories which can be characterized as "reductionist" in one sense or another. Reductionist learning theories take many forms and it will be impossible to examine the many varieties here. However, they all seem to share one overriding feature despite the varying interpretations of this feature. This feature is the desire to "get rid of" mysterious talk of mental states and processes in favor of good, solid, publicly observable, empirical facts. It has long been a puzzlement to philosophers, as well as others, just how certain "mentalistic" concepts in our vocabulary function. The primary examples of such concepts are 'thinks', 'believes', 'knows', 'intends', 'desires', 'wants', 'is aware of', etc. Furthermore these concepts seem distinctively "human" in the sense that they, unlike physical predicates such as 'weighs', 'is colored', 'occupies such and such a spatio-temporal location', and 'moves through space in such and such ways', seem to be applicable only to persons and sometimes to animals in lesser degrees. (I ignore until Chapter VI the ascription of these mentalistic predicates to machines.) There can be no doubting the centrality to psychology and philosophy of these mentalistic concepts, and yet they have proved to be remarkably resistant to attempts to explain them in any satisfactory way. The history of

philosophy abounds with such abortive attempts.

Modern reductionist psychology, however, has noted another feature of these mentalistic terms. This is that these terms are somehow middlemen between things which happen in the environment to a person and what the person subsequently does in the environment. Further, these happenings and doings seem to be describable in good, old-fashioned, publicly observable, empirical terms. Thus since there seems to be no clear-cut acceptable explanation of the role of the mentalistic middlemen, perhaps we can simply cut them out of the picture and examine merely the happenings and doings. Or to use a metaphor which has been popular in psychology, these mysterious states and occurrences can, for the purposes of empirical psychology, be put in a "black box" and ignored, attention being directed to the inputs to the black box and the outputs from it. It must be noted that this is not necessarily an ontological claim that the black box is contentless, nor even a claim as to the possibility of defining the contents of the black box in terms of its inputs and outputs. It is merely a methodological principle which might prove helpful in certain cases or certain investigations; and, if it is not followed slavishly, no objection can be raised to it on the grounds I have been considering.

In fact, most present day behavioristic psychologists do not seem to misuse the "black box" principle.

They treat it simply as a sort of useful methodological principle and recognize its limitations. More and more, present day behavioristic psychologists are even delving into the black box itself. They seem to do this in two ways. There is current in psychology a distinction between "intervening variables" and "hypothetical constructs."[8] Very loosely both of these terms have to do with the black box. The distinction between them is that intervening variables seem to serve merely mathematical purposes linking various laws into some sort of logically coherent theory. On the other hand, hypothetical constructs seem to have a life of their own, i.e. their meaning is not exhausted by pointing to their context of use in psychological laws and equations.

To take each in turn, I have no objection to intervening variables as a methodological device. They are probably extremely useful. However, if their use is coupled with some sort of ontological statement claiming that all the black box problems can be thought of in terms of intervening variables, then I think that this approach becomes much more like the kind of extreme reductionism which I will discuss shortly.

I also have no objection to the notion of hypothetical constructs. This notion goes even farther and begins describing the contents of the black box in terms

---

[8]See, for example, E.R. Hilgard, "Intervening Variables, Hypothetical Constructs, Parameters, and Constants," The American Journal of Psychology, Vol. LXXI (1958), 238-246.

far richer than their role as parameters in output-input equations. I am quite sympathetic to such an approach. Again, however, there is a danger. If the hypothetical constructs come to be looked upon as somehow definable in terms of or reducible to the input-output data, then this approach also becomes very much like the extreme reductionism to which I shall now turn.

The extreme reductionists I am considering are not content with the black box as a merely heuristic methodological principle. They seem to desire to reduce the putative contents of the black box to a theory of inputs and outputs whether these inputs and outputs are defined in terms of gross physical behavior or the microscopic concepts of physiology and neurology or something in between. It is at this point that the remarks I made in Chapter II concerning dispositions become extremely important. It will be recalled that I there claimed that there are important differences between dispositional statements and laws. However, I also claimed that it was very easy to slide from dispositional statements into laws. It is all too easy to collect a number of instances of the ascription of a disposition on the basis of certain behavior and generalize away the essential reference to a particular person. Furthermore, it is easy to slide into believing that the laws underlying dispositional statements

can be spelled out in the same form and with the same concepts as were used in the dispositional statement itself. I suggest that it is precisely this sort of confusion which underlies many of the more extreme reductionistic attempts. If someone believes something, then he is disposed to act in certain ways. By spelling out these various actions in their appropriate circumstances and observing many people, it seems that we can eventually completely eliminate the notion of belief altogether and talk simply of certain kinds of publicly observable doings. At least so the story goes. It is obvious that my analyses are not "reductive" in the sense under consideration, and it is also apparent that reductionists might claim that as a result, my analyses are useless. It is in clarifying such a possible controversy that the criteria I have developed come into play.

Although it is clear from the rest of his book that Cronbach is no reductionist, he gives a "postulate" linking learning with behavior which can easily be used to illustrate the problem with which I am concerned. So let me suppose, with apologies to Mr. Cronbach, that the following is to be taken as a linking postulate in a proposed reduction of ordinary talk about learning to behavior theory. "More precisely, learning is shown by a change in behavior as a result of experience."[9] I take

---

[9]Cronbach, op. cit., p. 71.

it here that 'experience' is to cover the inputs to the black box and 'behavior' the outputs.

I shall first suppose that the statement is to be taken as a stipulation. What then is being stipulated? It might be that 'experience' and 'behavior' are being stipulated respectively to mean that which causes learning and that which is shown by it. But then it would seem that lacking any but the ordinary language meaning of 'learning' or perhaps that given by my analyses, we have not at all succeeded in "getting rid of" mentalistic talk. On the contrary, we would by the stipulation have gotten rid of empirical talk about behavior and experience in favor of mentalistic talk about learning. It is far more likely that 'learning' is stipulated to mean what 'experience' and 'behavior' ordinarily mean. Now supposing that suitable clarifications of 'experience' and 'behavior' can be found which avoid mentalistic talk, there is nothing wrong with such a stipulation. (It is important to notice that the unwanted mentalistic talk is very often implicit in a precise characterization of 'behavior' and 'experience' when the stipulated connection comes to be extended in ways I shall discuss below. Of course, there is nothing wrong with such implicit use of mentalistic concepts except that it defeats the purpose of the reduction. In what follows I shall assume unless otherwise noted that

clarifications of 'behavior' and 'experience' do not contain unwanted mentalistic terms.)

The mistake, if mistake there be, comes when such a stipulation is extended with no further ado to cover situations in which 'learning' has its ordinary meaning. The problem is that the ordinary meaning of 'learning' has intentional uses which by virtue of the stipulation and the clarification of 'behavior' and 'experience' the stipulated meaning of 'learning' does not possess. All I am saying is that the legitimacy of such an extension must be argued for, and such an argument would lead, it seems, to one of the other kinds of linking postulate than the stipulative one. Thus if it is proposed to extend the above stipulated definition of 'learning' to cover ordinary cases of learning, some sort of argument concerning the equivalence of the ordinary uses of 'learning' with the uses of 'behavior' and 'experience' must be produced. All too many reductionist psychologists fail to recognize this simple logical point and all too many anti-reductionists are too hasty in assuming that psychologists are automatically attempting an unjustified extension whenever they make such a stipulation.

Let me now turn to the two other kinds of linking postulates. Unfortunately, it is often impossible to determine when a reductionist is putting forth his linking

postulates as empirical hypotheses and when he is making synonymy claims. In general, however, if the reductionist seems to take empirical evidence as justification for the linking assumption, then he means it to be an empirical hypothesis. On the other hand if he argues from the meanings of the word, he can be taken as asserting a synonymy claim.

Let me suppose that the statement is to be interpreted as a synonymy claim. It seems then that the anti-reductionist can quite plausibly argue that the proposed identification of meaning fails because 'learning' has intentional uses which the proposed definition lacks.

For example, contrast the following two descriptions of a basketball situation--the first in terms of learning and the second in terms of 'experience' and 'behavior':

(i) John deflected the pass because he had learned to keep his hands up.

(ii) John has in the past been on a basketball court. Occasionally the ball comes to his hands. Sometimes it leaves them. Sometimes the ball strikes John's hands and sometimes it doesn't, even though it would have struck his hands if they had been extended. Periodically, sounds issue from the coach's mouth which sound like "John, keep your hands up!" This time as the ball came flying

near John, his hands were extended and the ball bounced off of them.

Of course (ii) is horribly inadequate and it may be that in trying to characterize the coach's instructions I have used mentalistic terms. If I have, then the account would have to be expanded even more. But the difference between (i) and (ii) is not just that (i) is so much more efficient; it is that (i) refers to John as engaged in a purposeful activity, deflecting a pass, intending to deflect the pass, and doing so as a result of his learning. (ii) on the other hand might be describing a situation in which a ball accidentally bounced off John's hands. It says nothing about his learning or his intentions in making use of what he has learned. In addition it is very hard to see how (ii) could be used to describe John's intentions without allowing the very mentalistic concepts the reductionist wishes to avoid, for it is in general the case that mentalistic concepts are the ones which have the intentional uses.

Further the anti-reductionist may claim that we often use 'learning' to refer to situations where there is no observable behavior change whatsoever. A student may as a result of his experience of reading and studying a proof in Euclidean geometry learn that the theorem is proved in such and such a way. However, he may never be

called upon to prove that theorem. But if learning means a change of behavior, then we would have to say the student has not learned the proof; which may be patently false since he could have produced it if asked. It seems at the very least that the clarification of 'behavior' must take account of counterfactual exercises of learning which never actually result in any behavior change. Moreover, it will do no good for the reductionist to shift from grossly observable behavior to 'behavior' which is somehow described in neuro-physiological terms. It is patently false that 'learning' could mean that sort of behavior, for Euclid learned the proof, and understood what 'learning' meant although he was almost certainly totally unaware of neurophysiology. If the postulate is indeed a meaning claim, then the reductionist cannot ignore such arguments whether meaning is identical with use or only strongly connected with it.

There seem to be two possibilities open to the reductionist. He may claim that the "intentional" use of learning is not central to its meaning. But this requires an argument which is almost never found in psychological writings, probably because the claim is so obviously false. On the other hand if meaning is not identical with use, he may accept the fact that the intentional use is central to learning, and further accept the refutation of the

particular reduction he has proposed while maintaining that some other definition might be acceptable. Furthermore the fact that any acceptable definition of 'learning' must also have intentional uses may help guide his search for a new definition. For example, he might try to formulate the definition in terms of 'doing', 'saying', and 'perceiving'. These terms do have intentional uses and it is perhaps not immediately apparent that they involve mentalistic concepts, although plausible arguments have been advanced by anti-reductionists that they implicitly do make use of mentalistic concepts. The controversy would then move to the definitions and uses of these terms--the reductionist claiming no mentalistic concepts are involved and the anti-reductionist claiming that they are. What I am pointing out here is that the anti-reductionist cannot in general claim his arguments to be conclusive against any form of a proposed reduction although they may conclusively refute particular examples of it.

As far as I can tell there is no successful reduction of learning to behavior theory at present of the type which claims the linking postulates are true in virtue of the meanings of the words involved. Furthermore the anti-reductionist arguments are of such a nature that it is extremely difficult to see any reasonable way in which the reductionist might turn. However, lacking a Gödel

theorem for learning theory, it seems just barely logically possible that the reductionist might be able to yet come up with an acceptable definition of 'learning' which does not make use of terms he is unwilling to accept.

By far the most plausible thing for the reductionist to do, however, is to take the quotation as the beginning of an empirical hypothesis linking learning with experience and behavior. It might be objected at this point that the reductionist need not link learning with behavior and experience at all. Rather, following the heuristic principle of ignoring the black box altogether, he might simply seek laws connecting experience (input) to behavior (output). First, most present day behaviorists with a few notable exceptions do not believe that this can be done. They make free use in their theories of "intervening variables" or "hypothetical constructs" to connect the variables of stimulation (experience) with the variables of response (behavior). Of course they claim these intervening variables are unobsrvable, but they do feel a need for them.

Part of the reason for this use of intervening variables is that hypotheses using only experience and behavior are simply not in general confirmed except in extremely trivial cases. The most that such psychologists seem now to hope for are probabilistic laws connecting

stimuli and responses. But this raises another problem. Granted that statistical correlations can be made between certain experiences and behaviors, the question remains with these correlations whether they are really laws which enable the psychologist to explain and predict or whether they are merely generalizations which are not lawlike such as "Most of the coins in my pocket now are silver."[10]

The reason this question arises is that it seems prima facie plausible to argue that these generalizations are not lawlike because people can decide to act in any way they wish. That is, there seems to be a break in the chain of causes. The chain may lead perhaps from experience to belief and emotion, and from decision to action (behavior), but not from belief and emotion to decision. We seem to be able to choose to do what we want. Furthermore the fact that people are able to falsify such statistical generalizations if they know them and decide to act contrary to them seems to support the claim that the generalizations are not lawlike.

I cannot enter here on a detailed discussion of the problems of free will and determinism; it is sufficient for my purposes that the claim that statistical generalizations connecting experience with behavior are not

---

[10]Goodman, op. cit.

lawlike is universally rejected by reductionist psychologists of the kind I am considering, and, I believe, rightly so. But how do they do this? There seem to be two not unconnected ways. First, they sometimes claim that the generalizations are lawlike if the subjects do not know the generalizations and do not intend to falsify them, and are sincere in one way or another in their responses. But such an assumption quite obviously makes use of some of the contents of the black box, and thus the claim that the statistical correlations are lawlike rests in the end on considerations of what is in the black box. To argue that the connections between experience and behavior are lawlike is not to ignore the black box, but to make essential use of it.

Nevertheless, the more deterministic psychologists still claim to be able to do away with the mysterious mentalistic predicates even though they admit that the contents of the black box must be considered if they are to justify the claim that the statistical generalizations connecting experience and behavior are indeed lawlike. The most plausible way that they do this is by claiming that as a matter of empirical fact, 'learning' something is to be identified with certain brain processes, usually described in neuro-physiological terms. This sort of supposition together with some form of determinism provides

the guarantee that the statistical generalizations are law-like while enabling the psychologists to use mentalistic terms as intervening variables without specifying further just how these are eventually to be gotten rid of. That is the job of the neuro-physiologist. In any case there is no ignoring the contents of the black box; rather the psychologist claims that the mentalistic terms used to describe the contents can be gotten rid of by empirical laws connecting them to brain states and processes.

I have now reached what seems to me to be the most plausible reductionistic position concerning learning. What this position states is that 'learning' can be identified by the usual criteria we have always used--change of behavior as a result of experience--where 'behavior' and 'experience' can now include mentalistic terms in their clarifications. 'Behavior' also includes most importantly verbal reports of the contents of consciousness. There is no "reduction" here, that comes later; rather behavior and experience are the bases on which we ascribe mental predicates to people. Furthermore, it is clear that these criteria for the ascription of mental predicates are independent of the criteria for determining the brain states of any person. Empirical hypotheses can now be put forward linking the existence of the mental states described by the mental predicates to certain brain states

and these can be tested empirically.

Suppose, for example, that a certain brain state is linked by hypothesis with having learned the multiplication tables (by rote or with understanding). Suppose we then observe this brain state and ask the person to recite the tables or perform other tasks which we would take to be indications of his having learned them and he cannot do these tasks. Then it seems clear that this is a piece of disconfirming evidence for the hypothesis. A different hypothesis might be confirmed. What seems impossible is that any a priori argument could show that we never could get any laws of this form, for the laws are empirical ones.

But suppose that such laws were established, and suppose further that other requirements for a successful reduction such as theoretical simplicity, predictive power, and explanatory force are also satisfied. There seems to be no reason for anyone to deny that mentalistic concepts had been reduced to neuro-physiology. On the other hand it is not at all clear that the reduction will have gotten rid of a way of talking. Of course, very likely the neuro-physiological correlates of the mentalistic concepts will be extremely complex and we may have to retain the old way of talking for simplicity in communication. But this may not be the only reason for retaining the old ways of

talking. The situation may be analogous to that of the reduction of 'temperature' to molecular energy and not to that of the reduction of 'phlogiston' to classical thermodynamics. It may be possible to argue that talk in terms of "purposes," "beliefs," "intentions," "desires," "knowledge," "learning," is so basic to our way of looking at the world that we could not make sense of doing away with this way of talking even though the reduction has been successful. Such an argument, if sound, would have to show the extreme interconnectedness of this mentalistic vocabulary with our total conceptual scheme and would have to show that because of the "essential" uses of this vocabulary (essential to what we can now imagine) we would have but <u>one</u> subject matter but <u>two</u> essential vocabularies for describing it. At the very least when such a wonderful time arrives, if it ever does, the reductionist will not be able to rest on his laurels, impressive as they are. He will have to argue in terms of <u>meaning</u> and <u>use</u> for our mentalistic vocabulary's going the way of 'phlogiston'. And his opponent will have to argue in the same terms for keeping it on a par with 'temperature' despite successful reductions in all these cases.

At the very least I think I have shown in this chapter that the analyses I have given in terms of mentalistic concepts cannot, in the present state of the art,

nor even in the foreseeable future, be objected to on the grounds that they are not behavioristically reductionistic. I leave the question of any future reduction open, although I deny any present one is successful. 'Learning' is a "that which." It is that which is caused by paying attention to certain experiences in certain ways and that which enables us to plan and act in coping with the world. Whatever it may ultimately be, it is now not only legitimate, but useful, to analyze it in mentalistic terms.

CHAPTER V

AN EXAMINATION OF SELECTED EXPERIMENTS IN PSYCHOLOGICAL LEARNING THEORY

In this chapter I want to take a look at several experiments in psychological learning theory to see just how my analyses can help in understanding them, if at all. I hope to indicate some ways in which my analyses may be related to empirical work in psychology. This, I believe, will show that the sort of conceptual analysis with which I have been concerned is an aid to empirical work in learning theory; for if my analyses are to be anything more than mere verbal exercises, they must be anchored, at some point, to the empirical findings of learning theory.

1. A Simple Rote Learning Experiment

The first point which strikes one upon even a cursory glance at the experimental literature is that psychologists seem to identify rote learning with associationist psychology and principles. Something is rote learned if and only if it can be satisfactorily explained in terms of the principles of association between stimulus and response. Now without going into a detailed

history of associationist psychology, it seems evident enough that there is still some controversy over whether any principles of association can explain an intentional act of even a simple kind.

To take a familiar example, it is not immediately clear that Pavlov's dogs intend to salivate when they hear the bell or whether this "conditioned response" is the result of an association which would occur whether or not the dog intended to salivate or even if he intended not to salivate. Furthermore, this question of intention is not, I think, totally subsumable under the general question of whether to ascribe any mental attributes to animals at all. Rather, I believe, the same question could be asked of a human being. A human subject is conditioned to make a certain response upon the presentation of a certain stimulus. The question is whether the principles of associationist psychology can explain the subject's intending to make the response, or whether they can only explain the response disregarding the subject's intentions, and especially disregarding the possibility that he may intend not to make the response.

This question is, of course, crucial for me, since I have throughout been giving my analyses of learning in terms of acquisitions by the learner which he (the learner) can do with what he wills. In fact the element of

intention to perform what has been learned was what distinguished rote learning from mere conditioning (Chapters I and II). If the principles of associationist psychology can take care of this element of intention, then they very likely are studying rote learning. If they cannot, then they are studying a form of conditioned behavior which may share certain elements with rote learning, but which cannot be called "learning" except by stipulation, and the dangers of this have been discussed in detail in the last chapter.

Of course, I make no claims as to how "intention" must be taken care of by psychology. I make no anti-reductionist claims here. It simply seems to me that I have analyzed a very important concept which I believe with some justification can be called rote learning and that this concept differs from mere conditioning in that what has been learned is in a real sense at the subject's disposal, as opposed to conditioning, where what has been acquired cannot in general be performed at will. "Intention" is a "that which" which, of course, requires empirical specification, but which nevertheless can be used in a conceptual analysis to mark the logical differences between concepts.

The second striking point about psychological experiments with rote learning is the great deal of prior and even concomitant learning with understanding that the

experimenters assume. There must have been an enormous amount of 'learning that' and 'learning how' for the subject even to understand the usual verbal instructions given him in the experiment. Furthermore, it seems he almost invariably learns with understanding that the experimenter wants him to do, or try to do, such and such things. Even in the cases where the experimenter uses some sort of "deception" to hide just what the experimental purpose is, the subject learns with understanding that his task is such and such, even if certain other things which he believes he has learned that the experimenter wants are not actually so. And in these cases he most probably has learned with understanding in the derivative sense that the experimenter wants to test such and such a thing. And if he has an honest experimenter, he probably later learns with full understanding what the real purpose of the experiment was.

But even if these features are ignored, it is not altogether clear that typical "rote learning" experiments are really testing only rote learning and not some minimal form of learning with understanding as well. Let me now turn to a very simple and well accepted experiment which illustrates some of the remarks I have been making.[1] The

---

[1] John A. McGeoch, "The Influence of Associative Value upon the Difficulty of Nonsense-Syllable Lists," Journal of Genetic Psychology, Vol. 37 (1930), 421-426.

experimental procedure was quite simple. Two 10-syllable lists were constructed from each of six groups of syllables which had varying associative values determined by another method. Thirty-six college sophomores were given the lists to learn in varying orders and under varying time conditions. Each list was presented as a whole and learning was tested by immediate recall. The results show a definite, although not wholly uniform, tendency for more syllables to be learned as the associative value increases. Furthermore, lists of three letter words were learned most easily of all with almost perfect scores being recorded.

To begin with, this experiment is almost surely testing responses which are under the intentional control of the subject. When he is asked to recite a list, he almost surely does so intentionally and, furthermore, it seems clear that he could frustrate the experiment by simply deciding to answer incorrectly or not at all. The sort of experimental evidence which would show that this was only a conditioned response and not something which the subject learned might be something like the following. The subject is taught to give the response upon the presentation of a certain cue and then told not to give the response upon the next presentation of the cue. If he consistently does give the response anyway, and the experimenter is justified in believing that the subject did not

intend to give it, me might conclude that the subject had been conditioned.

It is interesting to note that here conditioning seems to be something more than mere rote learning, since ordinarily we would not expect such a conditioned response in human subjects without a great deal more and intensive conditioning than rote learning would seem to require. However, this does not affect the logical analyses of rote learning as being connected with the intentional exercise of what has been learned. It may indeed be necessary in such a situation to go through a great deal more conditioning to eliminate the operative intention of the subject, but this empirical requirement is immaterial to the logic of the concepts involved.

On the other hand, it is not necessary to go through rote learning before reaching conditioned responses. It has been shown that subjects involved in lengthy free association tests can be conditioned by nods of the experimenter's head to give many more plurals than would be expected by chance alone. Surely the subjects have not learned to do this even by rote, but nevertheless they have been conditioned to it. Of course, if they are told about this conditioning, they may intentionally eliminate the plurals, thus showing the intentional factor which is present in learning and absent in conditioning.

The major result of the experiment--the greater ease of memorization of words and word-like syllables--can very probably be explained on associationist principles. Obviously there is a greater association value with words which have already been learned in other contexts than with nonsense syllables. Metaphorically, the memory traces for the words are already there and can be utilized more easily. However, there is probably another operative factor in the greater relative ease with which words as opposed to nonsense syllables are learned. In order to see this factor, let me examine a hypothetical subject who has learned one of the lists.

Suppose a subject has learned a list composed of nonsense syllables of low associative value. How are we to describe this in terms of my analyses? Has he learned by rote to recite the list? Has he learned that the list is composed of a, b, c, . . .? Has he learned the latter with understanding? There seems little reason to deny that the subject has learned the list by rote. He has been in a series of trials which caused the probability to be near one that he would recite the list if he were now asked for it. Furthermore the subject's <u>performance</u> might also be rote. (Recall the distinction in Chapter 3, Section 1, between <u>learning</u> by rote and <u>performing</u> by rote.)

On the other hand the subject might very well also have learned that the list is composed of a, b, c, ... In fact it seems very unlikely in any verbal rote learning task such as this that 'learning that' does not also take place. The logical distinction between the two (if the 'learning that' is not 'learning with understanding that') would seem to be that belief is a necessary condition of 'learning that' but not of 'learning by rote to'. In the particular experiment just described the very procedure would make it highly unlikely that belief would not also occur. However, it seems just barely possible to distinguish empirically the two situations even here. The crucial thing would be to determine a situation in which the subject could recite the list and yet give evidence of not believing that the list is composed of a, b, c, . . . This might possibly occur if the subject could recite the list given the instructions "Recite list 3," but could not do it if the instructions were varied to any extent or if he were asked for the list at an unexpected time (assuming he had not forgotten). This would indicate a rote learning, but not a 'learning that', because belief that the list is composed of a, b, c, . . . would tend to be of wider applicability than an ability to recite the list in a narrow situation. Thus the subject's failures in wider situations might show that although he had rote learned

the list, he had not learned that the list was composed of a, b, c, ... Note that varying the instructions, e.g "What do you believe is in list 3?", can easily cause the subject to have the very belief in question. Thus this is a highly problematic situation at best.

However, if one moves from the verbal tasks which are so closely associated with belief to nonverbal tasks, the empirical distinction between 'rote learning' and 'learning that' is much clearer. Consider the case of a father teaching his son how to tie his four-in-hand. Suppose the father has learned by rote to tie a four-in-hand. Further he now does it by rote. He also probably learned that such and such was the appropriate way to tie it, but suppose he has by now forgotten this. The usual situation is that the father starts trying to tell his son how to tie the knot, finds he cannot do it, cannot even imagine how he does it, and must finally simply tie his own tie paying attention to what he is doing and by so doing, again learn that such and such is the correct procedure, even if he still cannot state the procedure very well. This seems to me a clear-cut case of someone having learned by rote to tie a four-in-hand, but not believing that such and such is the way to do it. He has learned it by rote and at best must relearn that such and such is the correct way to do it.

But now having admitted that 'rote learning' and 'learning that' almost invariably go hand in hand in verbal learning situations, let me ask if the subject might also have learned with <u>understanding</u> that the list is composed of a, b, c, ... I think that he very well might have. For the difference between 'learning that', and 'learning with understanding that' was seen to consist in the truth and justification of the latter as opposed to the former. Clearly the truth condition could be easily satisfied, and I claim that the subject could also have good reasons for believing that the list is composed of a, b, c, ... These reasons could include reinforcements from the experimenter, chances to check his recitation against the list, and even memory of a great number of past successes. Surprising as it may be, it is quite likely that a number of subjects in this rote learning experiment also learned with understanding, although as I shall show, the "understanding" is of a very minimal kind.

I think these considerations show two very important things. First, rote learning and learning with understanding are <u>very</u> closely connected. Not only must some things be rote learned in order to understand others, e.g. perhaps the multiplication tables to understand arithmetic, but mere rote tasks are very often accompanied by at least <u>some</u> learning with understanding. This last brings me to the

second important point. What is often thought of as a difference between rote learning and learning with understanding turns out to be rather a difference in the kind and scope of understanding involved.

To illustrate rather trivially this last point and also to round out the explanation of why the word lists are easier to memorize than the nonsense-syllable lists, let me again examine a hypothetical subject who has memorized both a word list and a nonsense-syllable list. Further suppose that he has learned that the lists are composed of a, b, c, ... To begin with, the empirical tests for distinguishing mere 'learning that' from 'learning with understanding that' would depend on the presence of good reasons for believing that one has the list right in the understanding case and the absence of these reasons in the other case. Thus if the experimenter withholds reinforcement and after the response does not allow the subject to check his results, and further is convinced that the subject does not otherwise have good reason to believe that he has the list right, then we might say that the subject has learned that the lists are composed of a, b, c, . . . but refuse to say that he has learned with understanding that the lists are composed of a, b, c, ...

On the other hand, if there is opportunity for reinforcement and the subject says something like, "Oh, I

know this is right; I remember its being right before and I've checked it rather recently" in response to the experimenter's question, "What makes you think you've got the list right?", then I think we could justifiably say that the subject has learned with understanding that the list is composed of a, b, c, ...

Another very important point is to be found in even this simple experiment. This is the role which reinforcement and opportunities to check the correctness of a response play in understanding. If my analyses are correct, and understanding is to be analyzed in terms of a person's reasons, then it is obvious that something very like reinforcement must be operative for the person to come to have these reasons. This point, along with the previous two, will be apparent time and again in what follows.

But to return to my subject who has learned with understanding both a word list and a nonsense-syllable list, it is clear that his reasons in the nonsense-syllable case are limited to the types which I have mentioned. On the other hand in the word list his reasons might also include such things as "The reason I know it's right is that I made up a little story which I remember very well and which contains the words in key positions," or "I know it's right because the first three words are connected with what I had for breakfast and the last seven with qualities in

people that I dislike." Clearly these sorts of reasons would be unavailable in the case of the nonsense-syllable list unless the syllables were very much like words. And sure enough, such lists fall in between total nonsense lists and word lists.

Let me repeat, I am not suggesting that the results of the experiment cannot be explained by associationist principles, nor even that the "understanding" involved here is of a very complex nature or differs very much from the word list to the nonsense list. I am simply trying to emphasize three important points which will be further examined in the following sections. First, pure rote learning seldom occurs without <u>some</u> learning with understanding. Second, apparent differences between rote learning and learning with understanding are really differences in the kind or scope of the understanding involved. And third, something very like reinforcement must occur if understanding is to occur on <u>any</u> level of sophistication. Moreover, all of these points are obviously related to my earlier discussion of learning a skill or discipline as opposed to learning more particularized things. It is the latter to which my analyses apply, and it seems to be the former which are being talked about by psychologists. In this section, as before, these larger units inherit their title to be called, e.g. rote, from the nature of the more

important or striking sub-parts.

2. Some Concept Formation Experiments by Bruner

In this section I want to look at some of the experiments described by Bruner in his book, A Study of Thinking.[2] In these experiments Bruner and his associates were primarily concerned with studying the various strategies used by subjects in attaining artificial concepts, comparing these strategies with certain logically "ideal" strategies, and noticing the results of varying certain features of presentation, cognitive strain, and difficulty, on the strategy adopted and the success obtained. I shall not be concerned with most of these details but rather with the overall experimental setup and certain of Bruner's general conclusions.

Very roughly most of the experiments were conducted with a group of cards upon which were imprinted various numbers and kinds of abstract figures in different colors and with different numbers of borders. In most of the experiments there were four attributes, each of which could vary in three different ways. These were as follows:

1. Shape of figure: circle, cross, square.
2. Color of figure: green, red, black.
3. Number of figures (all of same kind): one, two, three.

---

[2] J. S. Bruner, J. J. Goodnow, and G. A. Austin, A Study of Thinking (New York: John Wiley & Sons, Inc., 1956).

4. Number of borders: one, two, three.

The experimenter then chose a concept, e.g. "all red figures," or "yellow circles with one border," which the subject had to try to figure out. Usually the experiment began with the experimenter showing a card which exemplified either a positive or negative instance of the concept and telling the subject which it was (positive or negative). The subject then looked at other cards in various ways. Sometimes the subject chose them, sometimes they were presented randomly. Sometimes the subject had the whole array of cards before him in systematic order, sometimes not. Sometimes the subject could see as many cards as he wished, sometimes there was a limit on the number of cards he could examine and sometimes a time limit. In almost all cases the experimenter would indicate for each card in the sequence tested whether it was positive or negative and the subjects were always aware of the purposes of the experiment. There was no "deception." The subjects were always told they were to find the concept which the experimenter had in mind and in the case of "disjunctive" concepts, e.g. "green figures <u>or</u> three borders," they were also told that the concept <u>was</u> disjunctive and this notion was explained to them. In the other cases they were told that the concepts for which they were looking were of the usual "conjunctive" variety. The tests were performed almost

invariably on Harvard undergraduates.[3]

Two points stand out immediately in this description of the experimental procedure. First, the "concepts" learned in this experiment were highly artifical ones and bore little resemblance to the concepts of ordinary life. Although Bruner's notion of a concept as a well-defined class of objects may be formally equivalent to the denotation of ordinary predicates, the psychological differences seem a priori to be enormous. Surely some argument is needed to connect Bruner's concept formation with ordinary concept formation. Furthermore most ordinary concepts are tremendously vague and not at all as well-defined as the artifical concepts which Bruner used. For example, a car clearly is a "vehicle," but what of a child's wagon or tricycle or roller skates? Are these latter objects vehicles for some purposes and not for others, e.g. traffic laws? What about "defeasible" doncepts, i.e. concepts which do not pick out a class but give characteristics for exclusion? What is the role played in our conceptual scheme by the vaguenesses and ambiguities of ordinary concepts? To what extent is the learning of these vague concepts different from learning precise, simple, artificial concepts? These and a host of other questions need to be answered before Bruner's work can be taken to be more than a mere beginning in the direction of understanding concept

---

[3]The notions of "hypothesis testing" and "strategy" employed by Bruner in these experiments have been shown to

formation.

The second point is closely related with the first. Despite the artificiality of Bruner's concepts, it is clear that they make use of already quite familiar notions. Even the explanations given to the subjects of what the experimenter took a concept to be must have been readily grasped by the particular subjects. They already knew in a rough way what a "concept" was, and the explanations simply served to delimit these ideas for the purposes of the experimental study. What would Bruner have done had he had to teach his subjects ab initio what a "concept" was? What would have happened had the relevant attributes not been pointed out for the subjects or not been so apparent from the highly formal and abstract nature of the concepts to be learned?

The answer to this last question has been shown to be that the subjects would have had much, much more, difficulty.[4] None of these remarks are to be taken as criticisms of Bruner's work. Rather they serve to point out

---

be formally isomorphic to stimulus sampling theory. See P. Suppes, and R. C. Atkinson, Markov Learning Models for Multiperson Interactions (Stanford: Stanford University Press, 1960), p. 56.

[4]S.F. Osler and G. E. Trautman, "Concept Attainment: II. Effect of Stimulus Complexity upon Concept Attainment at Two Levels of Intelligence," Journal of Experimental Psychology, 62 (1961), 9-13. This experiment showed high intelligence subjects much more bothered in attaining the concept 'two' when the cards contained two pictures of natural objects than when they contained two abstract figures such as dots. Also see the following.

the dangers of extending his results to ordinary concept formation without answering some of these questions.

Let me now turn to a hypothetical subject who has been shown a positive card exemplifying a conjunctive concept and who has the time he wants and can choose any card he wants from a systematic array. One of the strategies commonly used by such subjects is what Bruner calls "conservative focussing." The subject uses the first card as a focus and systematically varies one attribute at a time in his subsequent choices. If he varies the number of figures and gets a positive response, he knows number is not relevant. If he varies color and gets a negative response, he knows the original "focus" color <u>is</u> relevant. And so he might proceed to establish the concept with minor excursions to check memory lapses or reconfirm some already obtained information.

Now omitting all the "incidental" learning which most of the subjects undoubtedly obtained, e.g. the purpose of the experiment, the nature of the cards, etc., I want to ask what sort of learning occurred in the main features of the experiment? To begin with it seems highly likely that in each case where a concept was finally attained the subject learned with understanding that the concept was such and such--and quite probably promptly forgot it. The "understanding" in such artificial cases as the one Bruner

considers is probably limited to such reasons as are peculiar only to the experiment. The subject learns with understanding that the concept is red circles. He believes it truly and his reasons are that the concept he is looking for is of this general type and the experimenter has told him that he has succeeded. Furthermore he is justified in believing the experimenter is truthful. He has learned with understanding that the concept is red circles. But surely <u>this</u> kind of understanding is not a very powerful one. It does not seem to be the one we are after when we say that someone understands something.

On the other hand the subject may have learned with understanding that the concept is red circles for entirely different reasons. This time he also believes it truly, but this time he knows the concept is red circles because he knows he has logically eliminated all the competing hypotheses by his strategy of conservative focussing. He knows that by varying one attribute at a time he can determine whether the attribute is relevant to the concept or not. For if he varies it, and the next instance is negative, he knows that the attribute was relevant. Further, he knows that the <u>one</u> change he made must be the one which changed the positive instance to a negative one. If an attribute is varied and the instance is positive, he knows that <u>that</u> attribute could not be

relevant, for the instance must still have all the relevant attributes to be positive. None of the relevant ones have been changed. This is the sort of understanding that one is intuitively after. It involves reasons which are much more powerful. The subject can use his developed strategy to solve similar problems whether there are truthful experimenters around or not, just as long as he can tell whether his instances are positive or negative. This requirement is one way in which he can come to have reasons.

But what about the learning of these strategies? Is understanding of them guaranteed simply by the nature of the experiment being one of concept formation? Is concept formation equivalent to "understanding" in some way? It seems to me that the answer to these last questions is clearly "No!" It seems obvious that one could "attain concepts" in this way without understanding what is going on at all. Let me try to show this by applying my learning analyses to the conservative focussing strategy.

The subject might use a conservative focussing strategy and become very successful at it. But I suggest that this might be due simply to his having learned by rote to vary one attribute at a time until the experimenter says he has got the concept. He may have learned

with understanding that the concept is such and such, but he has not learned with understanding that conservative focussing is a way to attain such concepts. The empirical tests to show this might include the subject's inability to state the strategy he uses or why it works and his inability to use it in similar situations without the artificial cards and the experimenter telling him when he is right.

Again the subject may have learned that varying one attribute at a time leads to the correct solution, but not understand this except for the reasons that it has worked in the past. This could be shown by the subject who could correctly state the strategy and use it in other quite similar situations but not be able to tell why it works or use it in situations where there is no one to tell him when the concept has finally been attained.

Finally the subject may have learned with "real" understanding that conservative focussing works. This subject could not only state the strategy and use it in widely varying situations, he would also know when he had eliminated all competing hypotheses and had the right one. He would no longer need the experimenter to tell him when he had the concept. The reasons connected with his understanding would be logically conclusive and he would know that they were.

It seems that all the subjects who used such a strategy would have learned how to attain the concepts, but the 'learnings that' connected with their common learning to use the strategy would be quite different. The first might have only learned that he always got the right answer by doing what he did, whatever that was. The second might have learned that he got the right answer by varying one attribute at a time. The third might have learned that he got the right answer by varying one attribute at a time because this would logically eliminate all the competing possibilities in the way I have sketched. All may have learned something with understanding, but the kind of understanding in each case would be vastly different.

Now it seems that the "interesting" or "good" kind of understanding is empirically shown by transfer. This kind of understanding is shown by the ability to state the strategy, to use it in other similar situations, and to be able to know the right answer without having to be told it. Transfer is indeed an empirical sign of understanding, but it is important to notice that it is so through the reasons which the subject has for his beliefs, in the way my analyses have indicated. The "best" kinds of reasons are those which pertain to the general relevant features of the task and only those features. These reasons would then be seen to be of the widest applicability.

The "worst" reasons are those which are limited to the specific task in question and would therefore have a very limited range of applicability. The trouble with the usual psychological use of transfer as a measure of understanding is that it is often far too narrowly conceived, e.g. the time savings in learning one task before another as opposed to learning each ab initio. Transfer is an empirical mark of understanding, but it must not be forgetten that it is so in virtue of its connection to the kinds of reasons involved in the understanding and these reasons can be manifested in a very great variety of ways.

Before leaving Bruner's series of experiments it will be instructive to consider the one time when he used more "concrete" cards. In this case the cards consisted of two figures, an adult and a child. Each figure could vary two ways in each of three attributes. Thus each figure could be of either sex, could be dressed in night or day clothes, and could be either holding out its hands to give a gift (adult) or receive one (child), or would be frowning with hands clasped behind his back (adult) or looking down with hands clasped behind his back (child). Clearly each card is evocative of a little story. For example a fully dressed woman frowning at a downcast boy with night clothes could be considered in terms of a mother berating her child for not getting up and getting dressed.

However, the concept formation tasks were still of the highly artificial type. They might be, for example, "all women giving," or "all children in night dress with a man." In other words the "stories" made no difference to the concepts being sought. Furthermore, the subjects were carefully informed of the type of concepts which were being sought.

As might be expected, however, the stories were used by the subjects in guiding their selection of cards to be tested. Certain attributes were changed much more often than chance would indicate. Others were changed less. The main conclusion reached by Bruner is probably correct: Meaningful material where the meaning is irrelevant to the concept being sought often interferes with correct solution. He also gives examples of this from the history of science where the French astronomer Lalande failed to discover Neptune despite making observations which might have led him to it. In accordance with the astronomy of his time he preferred the hypothesis that he had made an error in his observations rather than postulate a new planet. Bruner also mentions the failure of Curie and Savitch to discover atomic fission despite performing experiments which led others to do so later. All of this is clear enough. Interpreting situations in a familiar way when an unfamiliar way would be more

appropriate will often lead to erroneous or at least biased results.

However, Bruner sums up the experiment as follows:

> To attain concepts with materials that are meaningful and amenable to familiar forms of grouping leads to several difficulties. In the first place, the problem-solver is likely to fall back upon reasonable and familiar hypotheses about the possible groupings. In so doing, he may be led into a form of successive scanning: the strategy par excellence for going through a list of hypotheses. In the second place, the thematic material will, more readily than abstract material, lead certain attributes to have nonrational criteriality; the subject will "hang on" to these and will formulate hypotheses around them.[5]

Now although in this particular experiment some of the attributes did have "nonrational criteriality" since the subjects were told in effect that the "meaning" of the cards had nothing to do with the concepts being sought, it is far from evident that this would be so for less artificial learning tasks. What happened essentially in Bruner's experiment was that the subjects used "meaningful" attributes to construct hypotheses to test and were probably being "rational" in doing so since such attributes had very likely proved highly relevant in past ordinary learning tasks. Their behavior can be called non-rational only because they were implicitly given instructions in this case that the "meaningfulness" of the attributes was irrelevant, i.e. they were told that the concepts for which

---

[5]Bruner, op. cit., p. 111.

they were looking were in a sense artificial--just like the concepts in the prior experiments. Thus they had acquired, let me suppose, two opposing bits of learning: First, their long-time acquaintance with the relative importance of meaningful attributes, and second, their implicit instructions to disregard this in the present case. Is it any wonder that the more firmly established learning won out?

It seems that Bruner is here overlooking a very important distinction. It seems one must separate ordinary tasks of learning fairly well-accepted concepts from the tasks facing the problem-solver on the frontiers of knowledge. In the former case it would seem that far from being non-rational, a person who starts from the context of what is meaningful to him in a situation is not only being rational, but also learning in the most efficient possible way. And even in the latter case where the problem-solver is facing something entirely unfamiliar, it is a logical point that he must start with the features of the context which are meaningful to him or else he could never get off the ground. What this sort of problem-solver must be more aware of than the other is that there are quite a number of ways in which the meaningful attributes can be explained, and he should thus resist being led into the most familiar without considering the others. This is in

essence what happened to Lalande. He was more familiar with observational error--possibly due to his equipment--and so assumed his observations to be in error rather than investigate the alternative explanation of a new planet. This is also essentially Quine's point in talking about our conceptual scheme as a network which can be altered in innumerable ways to fit the facts.[6] Of course, one must be alert to the possibility of unfamiliar explanations, but this does not mean that one is being non-rational in solving problems by starting with what one knows and is most confident of.

3. A Concept Formation Experiment by Suppes and Ginsberg

In the preceding section I was especially concerned to bring out two important points. The first of these was that "pure" rote learning is really quite rare. In many so-called rote learning experiments there is a great deal of learning with understanding as well, although the kind of understanding can vary enormously. Connected with this point was the distinction between rote learning and rote performance. The latter may involve a great deal of learning with understanding although the former need not.

The second, not unrelated point, concerned the importance of distinguishing the various kinds of

---

[6]See previous discussion of this point in Chapter II, Section 4.

understanding involved. The first point was studied at some length in the preceding section. The point concerning the importance of the kind of understanding involved is illustrated quite well in an experiment performed by Patrick Suppes and Rose Ginsberg.[7]

The experimental procedure was as follows: The learning tasks were the equipollence of sets (same number of elements), the identity of sets (identical elements but unordered), and the identity of ordered sets (sets with identical elements in identical orders). Ninety-six first grade subjects were run in four groups of twenty-four each. In Group 1 the subjects first learned identity of sets for fifty-six trials and then equipollence for a further fifty-six trials. In Group 2 this order was reversed. In Group 3 the task was first identity of ordered sets and then identity of sets. In Group 4 identity of sets preceded identity of ordered sets. If we take the following symbolic representation (as used by Suppes in the experiment),

1. Ordered sets--O

---

[7]Patrick Suppes and Rose Ginsberg, "Experimental Studies of Mathematical Concept Formation in Young Children," Science Education, vol. 46, No. 3 (April, 1963), 230-240, and also reported in Patrick Suppes, "The Development of Mathematical Concepts in Children," Institute for Mathematical Studies in the Social Sciences, Stanford University, Psychology Series, Technical Report No. 63, February 25, 1964. It is the latter discussion with which I shall concern myself.

2. Identical sets--I

3. Equipollent sets--E

and $\overline{O}$, $\overline{I}$, and $\overline{E}$ for the absence of the property, we get the tasks as follows:

A

Group 1. Distinguish O and $I\overline{O}$ from $E\overline{I}$ and $\overline{E}$.

Group 2. Distinguish O, $I\overline{O}$, and $E\overline{I}$ from $\overline{E}$.

Group 3. Distinguish O from $I\overline{O}$, $E\overline{I}$, and $\overline{E}$.

Group 4. Distinguish O and $I\overline{O}$ from $E\overline{I}$ and $\overline{E}$.

B

Group 1. Distinguish O, $I\overline{O}$, and $E\overline{I}$ from $\overline{E}$.

Group 2. Distinguish O and $I\overline{O}$ from $E\overline{I}$ and $\overline{E}$.

Group 3. Distinguish O and $I\overline{O}$ from $E\overline{I}$ and $\overline{E}$.

Group 4. Distinguish O from $I\overline{O}$, $E\overline{I}$, and $\overline{E}$.

Of course the objective order of increasing generality is O, $I\overline{O}$, $E\overline{I}$, $\overline{E}$.

The sets in the stimulus display consisted of one, two, or three elements. On each trial two sets were presented and subjects were told merely to press one button when the sets were "the same" and the other button when they were "not the same." The subjects were required to make an overt correction response following each error. Furthermore, no stimulus display was ever repeated for an individual subject. This was done to rule out the possibility of the subject's learning the correct response by

rote. In the terms of my analyses the situation, S, varied enough from trial to trial to rule out rote learning. (See Chapter II, Section 6.)

It should be noted here that rote learning has not been logically ruled out. There seems to be the bare possibility that the stimulus displays or some other features of the trial were sufficiently alike in all correct cases that rote learning actually did occur. (All the other features are present.) However, the empirical possibility that rote learning occurred is extremely small and we are perfectly justified in assuming that the subjects did indeed come to learn <u>that</u> sets were "the same" whenever they had identical elements, for example. Therefore, they also believed it. This is a perfect example of evidence of such a belief despite the lack of any verbalization of it or even if it were badly verbalized. Of course, if a subject were asked why he thought the two sets were the same and he said something like "They both have people in them," then this would be evidence that he did not have the appropriate belief. Further tests would have to be made to either establish or refute the assumed correct belief. It would also be interesting to find out, if possible, why he said what he did.

Part of the experimental purpose was to investigate certain questions of transfer in the learning of

these obviously logically related concepts. The results were quite interesting. I shall not be concerned with the mathematical models used to interpret these results. These models are highly interesting in their own right, but a discussion of them would take me far afield. Rather, I shall merely indicate some of the results on transfer and difficulty of the various concepts.

First, it seems that the absence of any concept is more easily detected than its presence. This is explained by the experimenters as due to the "natural" dichotomy between O and $\bar{O}$. In all the other concepts the "presence" of the concept disagrees with this dichotomy. For example, both $I\bar{O}$ and $E\bar{I}$ show the presence of equipollence, but both of these are also examples of $\bar{O}$. On the other hand, $\bar{E}$ shows the absence of equipollence and also the absence of O thus agreeing with the natural dichotomy. It seems that the difficulty of a concept depends more on prior training and experience than on the concept itself, and this accords well with my analyses. The O versus $\bar{O}$ dichotomy is most "natural" simply because the subjects in previously learning what "the same" means have undoubtedly learned that two instances are the same if they have all relevant attributes in common. In the case of O versus $\bar{O}$ all the attributes except spatial location are the same. This is in contrast to the I vs. $\bar{I}$ or E vs. $\bar{E}$

case where order or kind of element can also differ. O examples accord more closely with the child's intuitive notion of "the same." A child who has learned with understanding that two O instances are "the same" has for his reasons his intuitive beliefs concerning what it takes to be the same. On the other hand these beliefs must be modified to a certain extent if he learns with understanding that two I or E instances are "the same." It is clear that the reasons involved in calling two instances "the same" can vary and with this variation comes differences in the kind of understanding.

Secondly, the experimenters found that in the objective order O, $I\bar{O}$, $E\bar{I}$, $\bar{E}$, the difficulty depended on where the "cut" was made in the order. Thus if the task were identity of sets, the cut would be between $I\bar{O}$ and $E\bar{I}$ with O and $I\bar{O}$ being the same (I), and $E\bar{I}$ and $\bar{E}$ different ($\bar{I}$). $I\bar{O}$ and $E\bar{I}$ turned out to be more difficult than O and $\bar{E}$. The nearer the cut the harder the class. This too accords well with my analyses. In terms of the child's reasons for believing two instances to be "the same" or "not the same," one would expect a priori O and $\bar{E}$ to be easier than $I\bar{O}$ and $E\bar{I}$, for the former two either have <u>all</u> attributes the same or almost none, while the latter two have some but not all attributes in common.

Perhaps the most interesting result occurs in the

transfer results from one task to another. Recall that the first task for Group 1 was to distinguish identity of sets, i.e. to distinguish O and $I\bar{O}$ from $E\bar{I}$ and $\bar{E}$. Their second task was equipollence, i.e. to distinguish O, $I\bar{O}$, and $E\bar{I}$ from $\bar{E}$. The observed results were a positive transfer for $I\bar{O}$ and a negative transfer (interference with learning) for $E\bar{I}$. The explanation seems clear. The natural dichotomy is O vs. $\bar{O}$. However, the first task gives reasons for calling $I\bar{O}$ "the same." This class carries over as "the same" to the second task, thus the positive transfer. On the other hand, the class $E\bar{I}$ is in the first task "not the same" and in the second task "the same"--hence the negative transfer.

Let me now look at these results from the point of view of my analyses. I shall assume that a hypothetical subject has learned with understanding at the end of the first task that two sets are "the same" if they have the same elements--identity of sets. What might his reasons for this belief be? He might remember that all those cases in which the two sets had the same elements in the same order were called "the same." He might also remember a case where one set had a dog and a boy and the other set had a boy and a dog in those orders respectively. This case was also called "the same" by the experimenter. On the other hand a set with a boy and dog was <u>not</u> called

"the same" as a set with a dog and a cat nor was a set with a boy called "the same" as a set with a boy and a cat. He now proceeds to his second task and suppose he also learns here with understanding that two sets are "the same" if they have the same number of elements. But his reasons here are different from his previous ones. He has had to learn in the course of trials in the second task that dog and boy is no longer "not the same" but "the same" as girl and cat. He understood the first task (identity of sets) in terms of reasons which did not also permit him to understand the equipollence of sets. He had to expand and modify his reasons to achieve the desired goal. Thus he had learned with understanding that identical sets have the same elements in the first task, but his reasons for this were different from his reasons in the second task where he had learned that equipollent sets have the same number of elements. After having learned both these things his understanding may be increased and his reasons for either belief probably are in terms that enable him to distinguish the two kinds of sameness. The importance of the kind of understanding in the two cases of learning with understanding is apparent.

In concluding this chapter I want to consider two general points that Suppes makes in his introduction to the above-mentioned paper, "The Development of Mathematical

Concepts in Children."[8] The first point that he makes concerns the banality of the usual distinctions made between understanding and rote skills. He argues that this banality is due primarily to the vaguenesses of the terms involved. It is obvious that I believe my analyses have helped to clear away these vaguenesses at least in part. But further they seem to give at least a partial philosophical clarification to the notions of stimulus transfer and generalization which Suppes uses to attack the banality. I have stressed the analysis of understanding in terms of the reasons which a person must have, but of even more importance is the fact that my analyses leave open, and even indicate, that there can be various kinds of understanding depending on the particular nature of the reasons. Thus there may be authoritarian reasons, logical coherence reasons, memory reasons, and theoretical reasons to name just a few. The "desirable" kind of understanding seems to involve the most general theoretical and logical reasons consistent with the person's ability to comprehend them. Finally it is an easy step to see the connections between the general theoretical reasons and increased stimulus transfer and generalization. Thus this latter psychological approach to evidence of understanding is seen to be naturally consistent with my analyses in terms of the kinds

---

[8] Ibid.

of understanding involved. Further a closer attention to the kind of understanding sought and the kinds which are possible in any given task will probably aid the psychologist's tasks in designing experiments and theory.

Another point which Suppes makes in connection with the banality of the distinction concerns the tendency of certain advocates of, for example, the new mathematics, to extend the term "rote" to situations which cannot be explained by simple associationist principles of psychology. I believe my analyses help clarify this situation in at least two ways. First I distinguish sharply between learning by rote and performing by rote. The former I have argued appears by itself very rarely and the latter may depend on a very high degree of previous learning with understanding. Thus I am on Suppes' side in believing that "rote" is far too freely used without specifying what is meant. Secondly, I have urged that there is a difference between learning with understanding a single sentence and learning with understanding a large body of material. I have argued that the latter, which seems to be what the new mathematics advocates are talking about, may indeed require the acquisition of a great number of rote skills possibly even learning them by rote. These rote skills are tied together with certain "singular" "learning with understanding that's." Again a clear understanding of what

is being called "rote" and the different kinds of understanding involved should aid discussion a good deal.

The second point Suppes is concerned with is the matter of the efficacy of the "discovery" method of teaching. Roughly, this method purports to allow the student to "discover" truths for himself as opposed to simply giving him material to be learned. Suppes objects that this method has never been formulated in any systematic way and that its advocates seem to resist any experimental tests by claiming that they are never of a sufficiently prolonged or general nature to really prove the method's value. It seems they always have a built-in objection to any empirical evidence which might tend to refute them. It seems to me that Suppes' objections are correct and that it is the intellectual responsibility of the advocates of discovery teaching to formulate their position so that it can at least begin to be tested. Nevertheless, it seems that my analyses might help to separate the wheat from the chaff in these discussions. If learning with understanding consists among other things in coming to have reasons of a certain sort, then I would venture the guess that the proponents of discovery have, perhaps without fully realizing it, gotten hold of this important feature. As I have already mentioned in Chapter III, the "reasons" which a person must have if he is to understand anything

must be comprehensible to him. It seems that the "discovery" method stresses this point by at least claiming to allow the child to "discover" such truths for himself. Implicit in this notion of "discovery" seems to be the idea that the child relates the newly discovered material with what he already knows, thus somehow logically guaranteeing that he will really have reasons.

On the other hand, as is so often the case when methodology itself becomes the guiding light, such people seem not to realize that there is no logical reason why other methods of teaching could not reach the same result and perhaps even more efficiently. It is even possible that the discovery method will not lead to the desired results of putting the reasons into the learner's own language, unless, as seems to be the case, 'discovery' is taken by these people as logically guaranteeing such acquisition of comprehensible reasons. But in that case they are not offering an experimental method, but merely calling attention to a logical feature of learning with understanding. But this is surely not what they claim to be doing. It seems entirely possible to me, given a sufficient knowledge of the general conceptual apparatus, terms, and modes of thinking of a group of people, that some lecturer could so devise a set of lectures using these terms that the resultant learning would be far superior to any "discovery"

method. I would suggest that the advocates of "discovery" teaching have dimly gotten ahold of a logically necessary condition of learning with understanding and then without further ado attempted to make this logical condition into a psychological method. Realizing the logical conditions is very desirable and may help in the construction and testing of methods, but this latter is a separate empirical question which requires a properly formulated empirical answer. This problem will continue to occupy me in the next chapter.

CHAPTER VI

APPLICATIONS AND CONCLUDING REMARKS

In the preceding chapters I have been concerned to emphasize, among other things, that there are various kinds and levels of understanding involved in the general rubric, "learning with understanding." For example, if a high school physics teacher says in a lecture that Newtonian mechanics is inadequate as a physical theory, then given certain assumptions concerning the general reliability of the teacher's pronouncements, the truth of what he says, and the student's belief in it, I would want to claim that the student may have learned with understanding that Newtonian mechanics is inadequate as a physical theory. But surely this kind of understanding is a very odd sort and not at all the kind that people seem to be after when they talk about learning something with "real" understanding. In this chapter I want to examine in a bit more detail just what the reasons must be like for a student to have learned something with "real" understanding.

A second point which I want to at least mention is the rising problem of machines that "think" and

"learn." I shall discuss this possibility and try to point out some conditions which must be satisfied if we are to allow that machines can learn.

1. Learning the Structure of a Subject with Understanding

The intuitive notion that one has concerning "real" understanding of a subject matter seems to involve the student's reasons being somehow directly connected with the subject matter itself and not with such apparently irrelevant things as some authority's telling the student something is the case, or the student's own vague intuitions about what ought to be the case. This rather vague notion can be sharpened somewhat by talking, as many now seem to be doing, about the structure of a subject. It seems that to learn something with real understanding, one must have as his reasons a grasp of the theoretical structure of the subject in question. This is, of course, closely connected with what it means to learn a subject matter or discipline with understanding, and I have already treated this problem in some detail in Chapter III, Section 2. Many of the ideas developed there are related to the present discussion, and I shall constantly be making use of them.

But what is the "structure" of a subject matter? Jerome Bruner gives us a clue when he says

> Grasping the structure of a subject is understanding it in a way that permits many other things to be related to it meaningfully. To learn structure, in short, is to learn how things are related.[1]

and,

> The teaching and learning of structure, rather than simply the mastery of facts and techniques, is at the center of the classic problem of transfer. There are many things that go into learning of this kind, not the least of which are supporting habits and skills that make possible the active use of the materials one has come to understand. If earlier learning is to render later learning easier, it must do so by providing a general picture in terms of which the relations between things encountered earlier and later are made as clear as possible.[2]

and further,

> The first object of any act of learning, over and beyond the pleasure it may give, is that it should serve us in the future. Learning should not only take us somewhere; it should allow us later to go further more easily.[3]

Several things are suggested by these excerpts. First, it seems that the learning of structure often involves learning certain subskills and a certain amount of "mere" facts. I have already argued that this is indeed true, and I shall not repeat the arguments here. It is sufficient to remark once again that learning a subject matter with understanding often involves learning <u>how</u> to do certain things and learning to <u>do</u> certain things in

---

[1]Jerome S. Bruner, <u>The Process of Education</u> (New York: Vintage Books, a division of Random House, 1960), p. 7.

[2]<u>Ibid</u>., p. 12.

[3]<u>Ibid</u>., p. 17.

certain situations in addition to learning with understanding that certain propositions of the subject matter are the case.

A second, and perhaps more important, point which is suggested by the quotations is, that of the propositions which express the truths of a subject matter, the most important for understanding are those which give the relations of the subject matter, i.e. the principles and laws of the subject. To learn the structure of a subject is thus to learn its principles.

Now I want to be clear at this point just what sorts of principles I am considering. In one sense, it could probably be justifiably said that every time anyone learns anything he always learns a principle; for this is just what makes learning different from merely hearing. For example, it might be said that if someone has learned that the Declaration of Independence was signed in 1776, then he has learned that whenever anyone asks him when the Declaration of Independence was signed, he should answer "1776." I admit that such "principles" are indeed involved in learning, but they are of the kind which relate the student to certain possible actions of his and are only indirectly connected with the subject matter he is learning. These are not the kind of principles I am considering.

Rather the principles I am talking about are those general statements, axioms, and empirical laws, which together with the definitions of the technical terms, compose what can be called the "theory" of the subject matter or discipline. These are the statements which enable one to relate the particular empirical facts of a discipline into a meaningful whole. In fact, one way of explaining any particular fact is to show how the statement of it follows deductively from a statement of certain general laws and boundary conditions.[4]

However, the learning of these laws can still take different forms and these different forms will reflect different degrees of understanding of the subject matter as a whole. For example, suppose that the associative, distributive, and commutative laws are part of the "structure" to be learned in learning how to solve algebraic equations. One might simply learn to transform such equations in accordance with these rules but not learn what these rules are nor how they are to be understood. Conceivably a student might be able to solve a great many equations in this manner, although he might be bothered by certain unfamiliar situations in which the rules do not have a straightforward application. Nevertheless, the student "understands" in some minimal sense how to solve algebraic equations, despite his not having

[4]See, for example, C. G. Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," Philosophy of

learned with understanding that the way one does this is by an application of these rules.

Next suppose that the student learns with understanding that the rules state such and such. Suppose, however, that his reasons for believing this are simply that his teacher has told him so and perhaps showed him how to apply the rules. In this case the student has, perhaps, a somewhat deeper understanding of how to solve algebraic equations. He not only manipulates the equations in accordance with the rules, but he can also state the rules and presumably apply them in relatively unfamiliar situations. However, his understanding of the "structure" of algebraic equations is still not as deep as it might be.

Finally, suppose that the student now turns to the foundations of arithmetic and begins to understand the basic structure of the real numbers in terms of sets, bounds of sets, ordering relations, and so on. I think that no one would hesitate in saying that such a student had a "real" grasp of the "structure" of algebraic equations and how the associative, commutative, and distributive laws work. And even here there might be a distinction between someone who understands these foundational matters and the professional logician who not only understands

---

Science, Vol. 15 (1948), pp. 135-175, reprinted in Feigl and Brodbeck, Readings in the Philosophy of Science (New York: Appleton-Century-Crofts, 1953).

them but can use them in discovering new truths.

The foregoing three cases can easily be seen in terms of concrete examples. Thus the first case might be the situation of a high school student in his first algebra class. The second might be that of a freshman in college majoring in engineering who had no desire to learn more than how to apply the rules. The last case could be a theoretical mathematician who has made it his life's work to understand the structure of algebraic equations.

What all this shows, I believe, is that simply appealing to "structure" does not take us very far in talking about learning with "real" understanding. Roughly speaking, knowing the structure of a subject matter may be a necessary condition of understanding, but it is certainly not sufficient since the structure can be approached in many different ways. It seems that there are quite a number of different kinds of reasons that we can use to justify our beliefs. And these different reasons lead to different kinds or levels of understanding. Thus hearing a person in a position of authority tell us something is one way in which we can learn something with understanding. Reading it in a book is another. Perceiving that it is so is yet another, while being able to place it in a theoretical framework is again another way in which one can learn something with understanding. And all of these ways, and

probably many more besides, lead to different kinds or levels of understanding.

These remarks also serve to show why a general analysis of having good reasons is so difficult to achieve without circularity. All of these different ways in which we justify our beliefs are prima facie "good enough." Of course, they can be contradicted by other of the ways and then some sort of weighing process seems to be needed to sort out what we are to believe with justification. That is, if our senses tell us one thing and a theory another, we must somehow decide which to believe if we are to achieve knowledge. But what I would like to suggest here is that while these various prima facie ways of justifying our beliefs are meant originally to help us understand what it is to have a rational belief, they themselves in the end rest upon our reconciling conflicts between them in a rational way. It seems that within certain bounds we <u>decide</u> in some sense what is to count as a justification of a belief. The notion of justification is essentially open-ended. If this is so, then a complete specification of what it is ultimately to understand anything is not possible. For particular inquiries at particular times and for particular people, the conditions for understanding are more or less fixed, and it is within each of these particular contexts that we can determine whether or not

something has been learned with understanding.

The preceding discussion has obvious and direct effects on questions of curriculum design. Even if we grant the importance of emphsizing the structure of any given subject matter, there still remains the question, and I think it is to a large extent an evaluative question, of just what sort of a grasp of the structure is to be required. Are we going to insist that high school algebra students learn the structure of equations in the same way and with the same understanding as the expert in the foundations of mathematics? I think not. Of course, both are studying the "same" structure, but the reasons they have for their beliefs about this structure will be vastly different, and, I think, rightly so. Value judgments on the kind of understanding required in any given subject at any given point will still have to be made.

This theoretical discussion is also related to the problem of teaching machines. Leaving aside the psychological problems of possible alienation and motivation, there seems to be no logical reason why machines cannot lead to learning with understanding. The reason for this is that the kind of understanding desired must be specified for given situations, and, once this is done, there seems to be no logical reason why machine programs cannot be designed to achieve the specified sort of understanding.

The critics of teaching machines cannot fall back on the argument that machines are not able to promote "real" understanding, because, if my arguments are correct, there is no such general thing. There are only particular kinds or levels of understanding which must be specified before any evaluations can be made as to whether understanding has been achieved or not.

But it might be claimed that my objections to structure, resting as they do on the notion of different kinds of understanding of structure, have been anticipated by the notion of a spiral curriculum.

> We begin with the hypothesis that any subject can be taught effectively in some intellectually honest form to any child at any stage of development.[5]

Roughly, the idea of a spiral curriculum is that of a curriculum which takes Bruner's hypothesis to heart. The structure of a subject is isolated and means of presenting it to students in various stages of development are devised. Instruction begins as early as possible and the curriculum spirals back upon itself several times, each time going into the subject more fully and with more depth. Essentially then, this is a decision as to what the various kinds or levels of understanding are which are desired at any point in the curriculum and an attempt to devise a program to produce these different levels of understanding.

---

[5]Bruner, op. cit., p. 33. See also pp. 13, 52-54.

Ideally a student could leave the curriculum at any point and still have a certain amount of understanding of the structure of that particular subject. In my example of teaching students how to solve algebraic equations, a student who wanted to be an artist might leave the curriculum after he had only learned to solve equations in accordance with the rules. An engineer might leave after learning how to solve the equations, and only a mathematician interested in foundations might follow the curriculum through to the deepest understanding of the structure of algebraic equations.

I could not be in more agreement with such a conception. If my analyses are right, then this sort of curriculum is the correct approach to learning with understanding on any level. However, noting these philosophical points does not guarantee that a suitable curriculum will be devised without a great deal more empirical work. Bruner seems to realize this, but he is not at all explicit as to just what the empirical work may involve.

For example, granted that the structure of a subject is what is to be learned with understanding, the best way of approaching this structure at various levels in the curriculum is not at all obvious. The structure of most scientific fields is usually conceived of as a hypothetico-deductive system. But surely one could not, without

empirical confirmation, simply present this system to the student in that form and believe that one had guaranteed understanding. Perhaps one should begin by allowing the students to try to discover this structure for themselves, gaining only an intuitive grasp of the formal properties of the system. This seems to be the idea of "discovery" teaching. Perhaps, however, it would be better first to try to teach the student some elementary properties of uninterpreted formal systems and only later get them to see how these systems can be given empirical significance. Perhaps some combination of these approaches would be best. Perhaps different approaches are needed at different stages. What should the role of actual empirical laboratory work be? How much methodology should be explicitly taught and how much discovered by the student? These and a host of other questions must be given empirical answers before the philosophically sound notion of teaching "structure" can be transformed into practical success.

In the humanistic studies where hypothetico-deductive systems do not exist, the structure of a subject is sometimes given in terms of the great themes which run through the subject. Again the question arises whether these themes should be taught first and then applied to concrete situations, or whether they should be developed and brought out through study of particular cases. Again

a great deal of empirical work is needed.

Furthermore, the empirical work is not simple. I think that any of these approaches could fit the conceptual requirements of understanding which I have given, but the problem is to make sure that each approach actually does do this. Any approach must take account of the stage of conceptual and cognitive development of the student, but all too often this is done only for the favored method of instruction. For example, suppose that one were teaching logic and had, what does not exist, a fairly complete picture of the conceptual scheme of a first grader. Suppose also that this scheme showed that there was very little grasp of formal operations. One might be tempted to conclude that the best way of teaching logical concepts to such students is simply to ignore the formalism and concentrate on inutitive discovery of logical notions through concrete examples. But this would be a mistake. It <u>might</u> be better, again starting from what the student actually <u>does</u> know, to try to develop an understanding of formal properties and later see how these are to be interpreted. Only adequate empirical research which does not bias the results a priori can decide the question. The only logical requirement is that <u>whatever</u> method is used, it must start with what the students do understand and thus can fit into their conceptual apparatus.

Furthermore, there is no reason not to try to change, by different prior instructions, the conceptual apparatus at any given age.

But even supposing that all these questions can be answered, there is still a great deal of ambiguity in the notion of "structure." Even among the best minds there is not universal agreement on what "the" structure of any given subject is, and this is true of the "exact" sciences as well as the inexact studies. For example, it is not at all clear whether the structure of psychology is to be understood as a search for merely descriptive generalizations of stimuli and responses as Skinner proposes, or whether a more complete theoretical framework is ultimately to be gained. To take another example from logic, there is still quite a bit of philosophical controversy as to whether the set-theoretical, truth-functional approach or the constructive, intuitionistic approach is the proper way to view "the" structure of logic. Surely the various interpretations of "the" structure of a subject must also be subjected to empirical test to see which one leads most easily to a complete understanding of the subject. And here "complete" understanding must include an appreciation of the differing ways of looking at "the" structure of the subject.

In the field of logic, experiments are now being

undertaken at Stanford with the help of Dr. Paul Lorenzen of Erlangen to begin to try to see whether an intuitionistic approach might in the long run produce better results than the truth functional approach. The experiments are only being begun and obviously this is a most complicated empirical task, but one which must be undertaken if we are ever to know the best way of teaching the structure of logic. Similar experiments need to be done in other fields. Perhaps these problems are not crucial at the beginning of the spiral, but they certainly are as one advances.

Another problem involved in the teaching and learning of the structure of a subject centers around the methodology of the subject. No matter what the method used to grasp the structure of a subject, that structure itself is, by its very nature, of a formal nature. This seems to be a conceptual truth, for the less easily one can formulate a subject in fairly precise terms, the less "structure" it seems to have. Now a more or less precise formulation seems to be necessary in order for a subject to be well understood. But once the basic structure has been learned, the worker on the frontiers of knowledge of that subject often does not use a precise formulation in order to guide further research. Indeed the further research might itself be the search for a more precise

formulation of a point which is grasped only intuitively.

Now it is this intuitive or insightful thinking of the research worker as compared with the analytic exposition of a well-defined area of knowledge that sometimes seems to be left out in discussions of learning with understanding. Can the methodology of discovering new truths be learned with understanding? I think that to a certain extent it can. There are, however, at least two obvious dangers here. The first concerns the nature of the understanding of the subject matter itself. It will be recalled that in Chapter II, Section 5, I discussed the example of a scientist who makes a good intuitive guess that something is the case. I claimed that although I might want to say that he knew that it was the case, I would want to deny that he had learned it with understanding unless he could fit his knowledge into the theoretical framework of his science and prove that it was correct, i.e. unless he <u>had</u> good reasons for his belief.

Now if this scientist consistently came up with good intuitive guesses, one might be tempted to say that he had learned with understanding how to discover new truths. I think this would be a mistake if the scientist could not give any reasons or tell how he made these guesses. He has learned <u>to</u> make good guesses, but he has not learned with understanding that he should do thus and

so to come up with these guesses; for his methodology has no explicit structure and thus could not be learned with understanding. It seems to me that the right thing to say about this case is that having learned a subject matter with understanding is a necessary condition of learning with understanding how consistently to discover new truths in the field, but it is not sufficient.

So understanding the methodology of a subject can be distinguished from understanding the subject. But this gives rise to the second problem. It will be recalled that in Chapter III, Section 4, I urged that there were very few decidable procedures for discovering new truths. (By "decidable" I mean a procedure which, if accurately followed, guarantees the finding of a solution if one exists.) Rather methodology seems in general to be a collection of heuristic procedures which may <u>aid</u> in finding new truths, but do not <u>guarantee</u> success. Now it is surely worthwhile to examine these heuristic principles in a theoretical way and try to give them a certain structure, and perhaps even convert some of them into decidable procedures. However, the danger is that one will forget that methodology is usually <u>only</u> heuristic if one concentrates on learning the theory of methodology. One can become a slave to methodology and come to believe, usually mistakenly, that if a solution cannot be found by following the

methodology, then one does not exist. Or one can reject solutions found by methods which cannot be put into the theory of the methodology. I have mentioned this danger in the previous chapter. In effect if one gives a structure to methodology, one is tempted to believe that this structure is itself decidable and complete, and anything which does not fit it is worthless.

Of course, a great deal of empirical research is needed here as well to see what effects an explicit methodology has on intuitive thinking. The important thing to remember is that although an understanding of methodology is not guaranteed by understanding the subject, neither is understanding the methodology a guarantee that new truths can be found only within the methodology.

2. Can Machines Learn?[6]

In this section I want very briefly to examine the relationship of my analyses to the question of whether machines can learn and in what ways. I want to make it quite clear that I am in no way offering any answer to this problem because I think in fact none can be given at the present. What I am concerned to do is to isolate some essential concepts in my analyses and try to see what

---

[6]I am indebted in this section to Dr. F. E. Yates of the Department of Physiology at Stanford University for several very helpful discussions. I am also grateful to him for having allowed me to read an unpublished paper of his entitled, "Men and Machines."

conditions would have to be satisfied in order for me to be willing to say that these concepts are applicable to machines.

I do not want to get involved in a semantical discussion here of just what it is for something to be a machine. This question may some day have to be answered before we can answer the question whether machines can learn and think. However, I think it will be clear enough for what follows if I simply take a machine to be something like modern day computers, servo-mechanisms, and self-guidance systems plus future developments of these along the presently indicated lines. I also want to exclude man from the reference of 'machine' although he may be a machine in the last analysis.

It is obvious that there are two key concepts in my analyses of learning which are crucial for deciding whether or not machines can learn. These are the concepts of "belief" and "intention." It is clear that "belief" is essential; for in my analysis of learning with understanding the person must not only believe what he has learned, but also he must believe the reasons for what he has learned and believe that they <u>are</u> reasons. Even in the case of rote learning the person usually believes that he is in the appropriate situation to exercise what he has learned whenever he is in such a situation, and in the

case of the rote learning of facts he must believe the statement of the fact. Furthermore, in both learning with understanding and rote learning any exercise of what has been learned must be intentional in the sense in which I am using 'learning'.

To take the case of "belief" first, it is not at all clear that present day machines can be said to "believe" the propositions and rules that are stored in their memory banks. If we consider belief in its role as a disposition to do certain things (not necessarily intentionally) then it seems that we would be tempted to say that machines do believe. For surely the output of machines is governed by their programs and the information fed into them. We can also say that they produce their outputs <u>because</u> of the programs and data which they have. Furthermore, deterministic machines do so in accordance with laws which can be said to <u>underlie</u> the ascription of beliefs to the machines. If it is objected that laws do not underlie beliefs in the human case and that beliefs are only tendencies to act in certain ways, then random machines can be constructed which also only "tend" to perform in certain ways. I think that it is the dispositional aspects of beliefs which lead us to say that machines can "believe" things and I see no logical reason to deny that actual machines, and certainly possible ones, do seem to possess

the dispositional aspect of beliefs.

Furthermore, the fact that "belief" in the human case provides us with opaque contexts seems no reason to suppose that machines cannot believe things. For just as in the human case, if the machine does not have the information that a particular singular expression refers to the same thing as another singular expression, then the machine cannot substitute one for the other any more than the human can. Belief is opaque, at least in the simple cases, for machines as well as for humans.

But it might be objected that when we come to the more complex cases of belief and other important aspects of it, machines do not believe after all. In addition to beliefs being dispositional, the statements expressing them must be "grasped," or "understood," or "entertained" by the supposed believer, and this, it could be argued, cannot be done by a machine. But what is the force of the 'cannot' here? Is it a logical cannot or merely an empirical one at the present state of the art? It seems to me that this question goes to the very heart of understanding what language is all about.

It will be useful at this point to recall my discussion in Chapter IV, Section 1, about the distinction in language between meaning and use. If, as has been argued by some, language is to be understood in terms of a

rigorous formal theory with precise formulations of the syntax, semantics, and transformation rules, if, in short, meaning is logically independent of use, then there seems to be every reason to suppose that to grasp the meaning of a statement in the language is simply to "know" the formulation of the language and all the rules of it. In this case there seems to be no reason in principle why machines could not be built and programmed so that such a language could be understood in this way.

Now it is surely obvious that no ordinary language has such a precise structure. However, this may be entirely accidental and due simply to man's failures to understand the essential nature of language. It can then be argued that ordinary language can be replaced by one which does have the requisite logical rigor and furthermore we would be better off if this were done; for we would not be misled by an imprecise language. I repeat, if this is the nature of language, then it seems perfectly plausible to admit that a machine could be constructed that could "grasp" a proposition expressed in this language and hence <u>believe</u> the proposition.

On the other hand, if, as I have suggested, meaning is not completely logically distinct from use, then another problem arises. It may be that much of ordinary language can be understood in terms of a precise systematic

language, but this may not be the whole story. It is because language has what precision it does that we can use it in the ways in which we do, and it is through a knowledge of these uses as well as the precise characterizations of linguistic units that we are ultimately to understand language.

For example, it is just because words have the meanings that they do that we can understand what a metaphor is. According to the rules of a formalized language, metaphors would be either meaningless or false and that would be an end of it; for it is clear that a pretty girl is not, strictly speaking, a melody. Nevertheless, metaphors are not meaningless and although they may be false, that is not the end of the matter. The situation seems to be that in virtue of their meanings, metaphors can be used in a language-using society to serve certain intentions that the members of the society may have, e.g. calling attention to certain similarities without actually claiming that every aspect is the same.

Moreover, it does not seem likely that these uses could ever be given a strict formulation. For if such a formulation were given, it would always seem to be possible for someone to use the formulation for yet other purposes. In fact this seems to me to be the case in ordinary language with lying. We already have a fairly precise set

of rules that tell us what language to use in making an assertion, and it is precisely by using these rules with the intention of deceiving someone that lying is possible.

If all this is correct, and if it is impossible to understand language completely except in the context of a society of language users, then two points arise concerning the applicability of "belief" to machines. The first is that the notion of belief cannot be extricated from the notion of intention. I shall return to this point below. The second point is that, at the very least, for machines to "believe" something, they must be members of a society of some sort with societal purposes, sanctions, rules, etc. This society could be among men and machines or it could be an intermachine society.[7]

Of course, the philosophical questions concerning the basic structure of language are so unsettled that we cannot easily decide the above questions concerning the relative merits of precise languages versus ordinary languages. In other words it is not clear whether or not grasping the meaning of a sentence in a language can be understood independently of the societal use of that language. If it can, then there seems to be no logical

---

[7]For an interesting discussion of a similar point and an excellent exposition of the general position that machines can think see, J.J.C. Smart, Philosophy and Scientific Realism (London: Routledge and Kegan Paul, 1963), especially Chapters V and VI.

reason why machines cannot be said to believe things. If "grasping the meaning of a statement" does need to be understood in terms of societal uses, then it is simply not clear whether machines can believe or not. In either case if machines are to be said to have learned things, the question of whether they can "intend" must be answered, and it is to this question which I now turn.

For all that the notion of "grasping the meaning of a statement" may not depend on a society of communicating beings, the notion of "intention" clearly does. Thus at least one condition of machines being able to learn in the full sense is that they be capable of being members of some sort of society. They must be able to formulate purposes, communicate these purposes, and find means of carrying them out. And this involves me at long last in the question of free will; for this question is at the heart of "intending." It seems obvious enough that no present day machine exhibits anything like free will. It is true that certain machines, for example, guided missiles, exhibit a certain kind of "purposive" behavior, but this behavior is clearly not of the machines own choosing. It will not be necessary for me to get involved in a detailed discussion of free will and determinism in order to suggest at least two ways in which it might be possible to attribute "free will" to machines. If these

ways are indeed possible, then this should show that there is no logical reason why we may not some day be willing to ascribe free will to machines.

One school of philosophy has long argued that there is essentially no inconsistency in the notions of free will and determinism. Following Aristotle this school characterizes a "free" action as one which is caused, but caused in a particular way--by the motives, desires, beliefs, etc., of the agent. Freedom on this view is simply the absence of constraints, and there seems to be no logical absurdity in supposing a machine could so act. Of course, free action is also action which is open to reason, persuasion, threats, and so on. But again if the machine were complex enough, there is no reason to suppose that we could not program it to recognize certain incoming data as reasons, threats, etc., and modify its action accordingly. A great deal of work would proabably have to be done on the logic of such concepts as "want," for example, in order to be able to program a machine to do something whenever it "wants" to, but again I see no logical difficulties.

Alternatively, this approach is strengthened by each advance in deterministic psychology. If, for example, a program of neuro-physiological reduction can be carried out, then we would have a neural correlate of what we call free will and could very probably program the same structure

into a computer. It would seem that psychology and computer programmers could work hand in hand on this approach, each supplying models for the other to try. I see no reason why there may not be a convergence some day.

The other approach takes free will to be somehow indeterministic, uncaused. But again there are no logical problems here. From the side of humans this indeterminacy need be nothing more than a random leak of stored information into certain plans to produce altered plans which are then evaluated as to merit by other plans. After all, we are born with a genetic "plan" which determines to some extent our development. Might not some of these genetic plans be altered by incoming data and others by random leaks of some sort?

From the side of machines it is certainly possible to construct non-deterministic machines, machines which act on the basis of random data. Furthermore, we can construct machines which could randomly leak some of their data and thus change parts of their programs. These changed programs could be evaluated by other programs and so on. If such a machine seemed to exhibit intentional, rational behavior, would we wish to deny that it had free will? If so, why?

Admittedly all of the foregoing is highly speculative. Certainly no machine at present is capable of

such behavior, but might this not be simply because we have been too timid in our design? The important point I hope to have brought out is that if machines are ever to be developed which can be said to "believe" and "intend" in a full-blooded sense, then these machines must be designed to operate as members of a communicating society with all the complex behavior which that entails. But I have found no logical reason for believing that this might not some day come to pass.

This whole problem of whether machines can learn can be approached from another direction than seeing what conditions would have to be satisfied for my analyses to apply. This approach makes use of what machines which have already been developed can, in fact, do. There are two recent developments in this regard which are especially worthy of notice. The first is that machines have been developed which can, in a certain sense, "learn from experience." The second point is that machines have been developed which can recognize certain "gestalt" properties.

By "learning from experience" I mean that machines have been built which have hierarchical programs. This means that the machine can keep track of its behavior on one level by means of a program on a higher level. The higher level program can then change the lower level program in accordance with some predetermined criteria

of merit. Furthermore, the hierarchy seems to be limited only by the complexity and size of the machine.

This development makes possible a certain amount of "strategy" and "inventiveness" by the machines. For example, certain heuristic strategies can be built in on the second level program which can examine the first level operations and change them according, perhaps, to a third level criteria of merit. Thus results can be obtained which were not predicted by the programmer even though the working of the machine may be deterministic and the human calculator would have obtained the same results following the same rules.

Another process which can be used with hierarchical programming is an inductive one. Inductive rules of inference can be programmed into the computer at higher levels and the machine can then observe its own behavior and inductively formulate the rules of this behavior. Of course, these inductive rules lack a mathematical precision, but then so do the intuitive rules by which humans work. Even in logic the precise rules of a formal language can be formulated only in the less precise notions of a meta-language. And if one converts the meta-language into object language, then <u>its</u> rules can only be formulated in terms of yet another imprecise meta-language. The computer is no worse off than human beings in being compelled

to start with something which is simply intuitively "understood" or taken for granted.

One upshot of all this is that machines can in principle be constructed which can show Gödel sentences to be true.[8] Furthermore, the machine can convert itself into a higher order machine by adding the Gödel sentence to the axioms of the theory and constructing another Gödel sentence. The process is limited only by the size of the machine.

But not only can an inductive machine be constructed, one can also build in a capacity to take occasional leaps on less than inductively adequate evidence. This could be done in a random way and the conclusions thus arrived at could be evaluated by the machine. Surely it is an empirical possibility that such a procedure would come up with as many useful conclusions as human intuiters do. And if the machines did do this, would not this be some evidence for saying that computers can make intuitive "leaps" on the basis of their past "learning?" I think that these developments show that in a certain sense machines can solve some very complex problems in a non-trivial way.

---

[8] Gödel sentences are statements which can be formulated in a theory but which can neither be proved nor disproved in that theory. However, they can be shown to be true by metatheoretical considerations. Furthermore, if a Gödel sentence is added to the theory as an axiom, it is possible to construct a new Gödel sentence in the new theory, and this process can be repeated an infinite number of times.

Surely this is at least one element in using what has been learned.

The second major machine development which I want to discuss is the recent work in the recognition of "gestalt" properties of symbols. Machines have been built which can recognize words and letters in whatever size or position they may appear in its scanning field. For example, 'e' can be recognized as an e whether it is large or small, backwards, upside down, tilted, etc. Now surely this is at least a promising beginning in the recognition of essences, e.g. "e-ness." A great many of the problems arising because of the previous necessity of restricting computers to arithmetical data can be solved by such machines.

Of course, it can be objected that the "essence" that the machines recognize in such cases is the essence of a physical symbol and not the essence of what the symbol means. And I think that this would be a correct thing to say. However, as I have already argued, at least a part of the meanings of symbols can be expressed in rather precise terms and these surely could be programmed into the machine for use in "grasping" the meaning of words.

But this brings me right back to the problem of the distinction between the use and meaning of linguistic entities. If a knowledge of use is required in addition

to knowledge of formal properties in order to grasp the meaning of language, then it would seem that machines must be brought into some sort of society and must be able to formulate and carry out their own purposes in this society before one can say that they actually learn.

Can machines learn? I simply do not know. The answer to the question must await further empirical developments in the field of computer design. It seems clear enough that present day machines cannot learn in the sense that full-fledged members of a society can learn. Nevertheless, some of them certainly exhibit behavior sufficiently similar to the behavior of persons who have learned something to enable us to say at least in an inverted commas sens of 'learn' that machines can "learn." Furthermore, I have tried to point out just what the force of the inverted commas might be. I think it refers to the problem of constructing a machine which can freely interact in a society. And I see no a priori reason why these inverted commas may not someday be removed from the statement, 'Machines can "learn"'.

At any rate, humans can learn, and it is to some of the problems of human learning that this dissertation has been addressed. Although I make no claims to have solved or even touched on all the problems connected with

learning, I hope at least to have made a beginning at understanding some of the logical requirements of certain kinds of learning. If I have removed even some of the conceptual confusion that all too often surrounds discussions of learning, I will have succeeded.

SELECTED BIBLIOGRAPHY

Books

Anscombe, G.E.M. *Intention*. Oxford: Basil Blackwell, 1957.

Aquinas, Thomas. *Summa Theologica*.

Armstrong, D.M. *Perception and the Physical World*. London: Routledge and Kegan Paul, 1961.

Austin, J.L. *How to Do Things with Words*. Cambridge, Mass.: Harvard University Press, 1962.

Ayer, A.J. *The Problem of Knowledge*. London: MacMillan, 1956.

Bruner, J.S. *The Process of Education*. New York: Vintage Books, a division of Random House, 1960.

Bruner, J.S., Goodnow, J.J., and Austin, G.A. *A Study of Thinking*. New York: John Wiley & Sons, Inc., 1956.

Carnap, R. *The Logical Foundations of Probability*. Chicago: The University of Chicago Press, 2d ed., 1962.

Chisholm, R.M. *Perceiving: A Philosophical Study*. Ithaca, N.Y.: Cornell University Press, 1957.

Collingwood, R.G. *The Idea of History*. Oxford: Clarendon Press, 1946.

Cronbach, L.J. *Educational Psychology*. New York: Harcourt, Brace, and World, Inc., 2d ed., 1963.

Dewey, John. *Democracy and Education*. New York: MacMillan 1916.

_____. *Experience and Education*. New York: MacMillan, 1938.

Dray, William. *Laws and Explanation in History*. London: Oxford University Press, 1960.

Feigl, H. and Brodbeck, M. (eds.). *Readings in the Philosophy of Science*. New York: Appleton-Century-Crofts, 1953.

Feigl, H. and Maxwell, G. (eds.). Current Issues in the Philosophy of Science. New York: Holt, Rinehart and Winston, 1961.

Feigl, H. and Sellars, W. (eds.). Readings in Philosophical Analysis. New York: Appleton-Century-Crofts, 1949.

Flew, A.G.N. (ed.). Essays in Conceptual Analysis. London: MacMillan, 1956.

Goodman, Nelson. Fact, Fiction, and Forecast. Cambridge, Mass.: Harvard University Press, 1955.

Hare, R.M. Freedom and Reason. Oxford: Oxford University Press, 1963.

_____. The Language of Morals. Oxford: Oxford at the Clarendon Press, 1961.

Hilgard, E.R. Theories of Learning. New York: Appleton-Century-Crofts, 1st ed., 1948.

_____ (ed.). Theories of Learning and Instruction, NSSE Yearbook LXIII, Part I, Chicago: National Society for the Study of Education, 1964.

Hintikka, Jaakko. Knowledge and Belief. Ithaca, N.Y.: Cornell University Press, 1957.

Melden, A.I. (ed.). Essays in Moral Philosophy. Seattle: University of Washington Press, 1958.

Nagel, Ernest. The Structure of Science. New York: Harcourt, Brace, and World, Inc., 1961.

O'Connor, D.J. An Introduction to the Philosophy of Education. London: Routledge and Kegan Paul, 1957.

Philipson, Morris (ed.). Automation: Implications for the Future. New York: Vintage Books, a division of Random House, 1962.

Plato. Theaetetus.

Popper, K.R. The Open Society and Its Enemies. Princeton: Princeton University Press, 1950.

Price, Kingsley. Education and Philosophical Thought. Boston: Allyn and Bacon, Inc., 1962.

Quine, W.V.O. From a Logical Point of View. Cambridge, Mass.: Harvard University Press, 1953.

_____. Word and Object. New York: The Technology Press of the Massachusetts Institute of Technology and John Wiley and Sons, Inc., 1960.

Ryle, Gilbert. The Concept of Mind. New York: Barnes and Noble, 1949.

Scheffler, Israel. The Language of Education. Springfield, Ill.: Charles C. Thomas, 1960.

_____ (ed.). Philosophy and Education. Boston: Allyn and Bacon, Inc., 1958.

Smart, J.J.C. Philosophy and Scientific Realism. London: Routledge and Kegan Paul, 1963.

Smith, B.O. and Ennis, R.H. (eds.). Language and Concepts in Education. Chicago: Rand McNally & Co., 1961.

Strawson, P.F. Individuals. London: Methuen, 1959.

Suppes, P. and Atkinson, R.C. Markov Learning Models for Multiperson Interactions. Stanford: Stanford University Press, 1960.

Wiener, Norbert. God & Golem, Inc. Cambridge, Mass.: The Massachusetts Institute of Technology Press, 1964.

Wittgenstein, L. Philosophical Investigations. Trans. by G.E.M. Anscombe. New York: MacMillan, 1953.

Articles

Aaron, R.I. "Feeling Sure," Proceedings of the Aristotelian Society, Supplementary Vol. XXX (1956), 1-13.

Ammerman, Robert. "A Note on Knowing That," Analysis, XVII (1956), 30-32.

Arner, Douglas. "On Knowing," Philosophical Review, LXVIII (1959), 84-92.

Austin, J.L. "Other Minds," Proceedings of the Aristotelian Society, Supplementary Vol. XX (1946), 148-187.

Black, Max. "Certainty and Empirical Statements--Discussion," Mind, LI (1942), 361-367.

Broudy, Harry S. "How Philosohical Can Philosophy of Education Be?", Journal of Philosophy, LII (1955), 612-622.

Chisholm, R.M. "The Logic of Knowing," Journal of Philosophy, LX (1963), 773-795.

Clark, Michael. "Knowledge and Grounds: A Comment on Mr. Gettier's Paper," Analysis, XXIV (1964), 46-48.

Cronbach, L.J. and Meehl, P.E. "Construct Validity in Psychological Tests," Psychological Bulletin, LII (1955), 281-302.

Gellner, Ernest. "Knowing How and Validity," Analysis, XII (1951), 26-35.

Gettier, E.L. "Is Justified True Belief Knowledge?", Analysis, XXIII (1963), 121-123.

Hartland-Swann, John. "'Knowing That'--A Reply to Mr. Ammerman," Analysis, XVII (1957), 69-71.

_____. "The Logical Status of 'Knowing That'", Analysis, XVI (1956), 111-115.

Hempel, C.G. and Oppenheim, P. "Studies in the Logic of Explanation," Philosophy of Science, XV (1948), 135-175.

Hilgard, E.R. "Intervening Variables, Hypothetical Constructs, Parameters, and Constants," The American Journal of Psychology, LXXI (1958), 238-246.

Levison, A.B. "Logic, Language, and Consistency in Tarski's Theory of Truth," Philosophy and Phenomenological Research, XXV (1965), 384-392.

Malcolm, N. "Certainty and Empirical Statements," Mind, LI (1942), 18-46.

_____. "Knowledge and Belief," Mind, LXI (1952), 178-189.

McGeoch, J.A. "The Influence of Associative Value upon the Difficulty of Nonsense-Syllable Lists," Journal of Genetic Psychology, XXXVII (1930), 421-426.

Osler, S.F. and Trautman, G.E. "Concept Attainment: II. Effect of Stimulus Complexity upon Concept Attainment at Two Levels of Intelligence," Journal of Experimental Psychology, LXII (1961), 9-13.

Price, H.H. "Some Consderations About Belief," Proceedings of the Aristotelian Society, XXXV (1934-1935), 229-252.

Roland, Jane. "On 'Knowing How' and 'Knowing That'," Philosophical Review, LXVII (1958), 379-388.

Ryle, Gilbert. "Knowing How and Knowing That," Proceedings of the Aristotelian Society, XLVI (1945-1946), 1-16.

Saunders, J.T. and Champawat, N. "Mr. Clark's Definition of Knowledge," Analysis, XXV (1964), 8-9.

Sosa, Ernest. "The Analysis of 'Knowledge that p'," Analysis, XXV (1964), 1-8.

Strawson, P.F. "On Referring," Mind, LIX (1950), 320-344.

Suppes, P. "The Development of Mathematical Concepts in Children," Institute for Mathematical Studies in the Social Sciences, Stanford University Psychology Series, Technical Report No. 63, February 25, 1964.

Suppes, P. and Ginsberg, R. "Experimental Studies of Mathematical Concept Formation in Young Children," Science Education, XLVI (1962), 23-34.

Tarski, A. "The Semantic Conception of Truth and the Foundations of Semantics," Philosophy and Phenomenological Research, IV-V (1944), 341-376.

_____. "Der Wahrheitsbegriff in den Formalisierten Sprachen," (German translation of a book in Polish, 1933), Studia Philosophica, I (1935), 261-405.

White, A.R. "On Claiming to Know," Philosophical Review, LXVI (1957), 180-192.

Woozley, A.D. "Knowing and Not Knowing," Proceedings of the Aristotelian Society, LIII (1953), 151-172.

Unpublished Sources

Yates, F.E. "Men and Machines," Stanford Universitty, unpublished paper.

Thirza H. Hibner
12661 Robleda Road
Los Altos Hills