PSYCHOLOGICAL REVIEW, 1948, *55*, 95-107

#13

ON A DISTINCTION BETWEEN HYPOTHETICAL CONSTRUCTS AND INTERVENING VARIABLES

BY KENNETH MACCORQUODALE AND PAUL E. MEEHL

University of Minnesota

As the thinking of behavior theorists has become more sophisticated and self-conscious, there has been considerable discussion of the value and logical status of so-called 'intervening variables.' Hull speaks of "symbolic constructs, intervening variables, or hypothetical entities" (5, p. 22) and deals with them in his theoretical discussion as being roughly equivalent notions. At least, his exposition does not distinguish among them explicitly. In his presidential address on behavior at a choice point, Tolman inserts one of Hull's serial conditioning diagrams (11, p. 13) between the independent variables (maintenance schedule, goal object, etc.) and the dependent variable ('behavior ratio') to illustrate his concept of the intervening variable. This would seem to imply that Tolman views his 'intervening variables' as of the same character as Hull's. In view of this, it is somewhat surprising to discover that Skinner apparently feels that his formulations have a close affinity to those of Tolman, but are basically dissimilar to those of Hull (10, p. 436, 437). In advocating a theoretical structure which is 'descriptive' and 'positivistic,' he suggests that the model chosen by Hull (Newtonian mechanics) is not the most suitable model for purposes of behavior theory; and in general is critical of the whole postulate-deductive approach.

Simultaneously with these trends, one can still observe among 'tough-minded' psychologists the use of words such as 'unobservable' and 'hypothetical' in an essentially derogatory manner, and an almost compulsive fear of passing beyond the direct colligation of observable data. 'Fictions' and 'hypothetical entities' are sometimes introduced into a discussion of theory with a degree of trepidation and apology quite unlike the freedom with which physicists talk about atoms, mesons, fields, and the like. There also seems to be a tendency to treat all hypothetical constructs as on the same footing merely because they are hypothetical; so that we find people arguing that if neutrons are admissible in physics, it must be admissible for us to talk about, *e.g.*, the damming up of libido and its reversion to earlier channels.

The view which theoretical psychologists take toward intervening variables and hypothetical constructs will of course profoundly influence the direction of theoretical thought. Furthermore, what *kinds* of hypothetical constructs we become accustomed to thinking about will have a considerable impact upon theory creation. The present paper aims to present what seems to us a major problem in the conceptualization of intervening variables, without claiming to offer a wholly satisfactory solution. Chiefly, it is our aim here to make a distinction between two subclasses of intervening variables, or we prefer to say, between 'intervening variables' and 'hypothetical constructs' which we feel is fundamental but is currently being neglected.

We shall begin with a common-sense distinction, and proceed later to formulations of this distinction which we hope will be more rigorous. Naively, it would seem that there is a difference in logical status between constructs which involve the hypothesization of an *entity*, *proc-*

Reprinted in Meehl, *Selected philosophical and methodological papers* (pp. 249-264; C. A. Anderson & K. Gunderson, Eds.). Minneapolis, MN: University of Minnesota Press, 1991.

ess, or *event* which is not itself observed, and constructs which do not involve such hypothesization. For example, Skinner's 'reflex reserve' is definable in terms of the total available responses without further conditioning, whereas Hull's 'afferent neural interaction' involves the notion of processes within the nervous system which presumably occur within the objective physical system and which, under suitable conditions, we might observe directly. To take examples from another science in which we psychologists may have less stake in the distinction, one might contrast the notion of 'resistance' in electricity to the notion of 'electron.' The resistance of a piece of wire is what Carnap has called a *dispositional concept*, and is defined by a special type of implication relation. When we say that the resistance of a wire is such-and-such, we mean that "so-and-so volts will give a current of so-and-so amperes." (For a more precise formulation of this see Carnap, 3, p. 440.) Resistance, in other words, is 'operational' in a very direct and primitive sense. The electron, on the other hand, is supposedly an *entity* of some sort. Statements about the electron are, to be sure, supported by means of observational sentences. Nevertheless, it is no longer maintained even by positivists that this set of supporting sentences exhaust the entire *meaning* of the sentences about the electron. Reichenbach, for example, distinguishes *abstracta* from *illata* (from Lat. *infero*). The latter are 'inferred things,' such as molecules, other people's minds, and so on. They are believed in on the basis of our impressions, but the sentences involving them, even those asserting their existence, are not reducible to sentences about impressions. This is the epistemological form, at rock bottom level, of the distinction we wish to make here.

The introduction of the word 'entity' in our discussion has served merely to indicate the distinction, but in any crucial case there could be dispute as to whether a stated hypothesis involved the positing of an entity. For instance, is Hull's 'habit strength' an entity or not? Is 'drive' an entity? Is 'super-ego'?

Previous analyses of this difference may enable us to give a somewhat more precise formulation. These two kinds of concepts are variously distinguished by writers on philosophy of science. Feigl (personal communication) refers to *analytic* versus *existential* hypotheses. Benjamin (1) distinguishes between *abstractive* and *hypothetical* methods. In the abstractive or analytic method we merely neglect certain features of experience and group phenomena by a restricted set of properties into classes; relations between such classes can then be discovered empirically, and nothing has been added to the observed in the process. The hypothetical method, on the other hand, relates experiences "by inventing a fictitious substance or process or idea, in terms of which the experiences can be expressed. A hypothesis, in brief, correlates observations by adding something to them, while abstraction achieves the same end by subtracting something" (1, p. 184).

This quotation suggests to us at least three ways of stating the distinction we have in mind. First, it may be pointed out that in the statement of a hypothetical construction, as distinguished from an abstractive one, there occur words (other than the construct name itself) which are not explicitly defined by (or reduced to) the empirical relations. Once having set up sentences (postulates) containing these hypothetical words, we can arrive by deduction at empirical sentences which can themselves be tested. But the words themselves are not defined directly by or reducible to these empirical facts. This is not true of abstractive concepts, such as resistance or solubility or, say, 'drive' as used by Skinner. (We may neglect wholly non-committal words such as *state*, which specify nothing except that the conditions are internal.)

A second apparent difference between abstractive and hypothetical concepts is in their logical relation to the facts, *i.e.*, the observation-sentences and empirical laws which are the basis for believing them. In the case of sentences containing only abstractive concepts, the truth of the empirical laws constitutes *both the necessary and sufficient conditions* for the truth of the abstractive sentences. For sentences involving hypothetical concepts, this is well known to be false. The empirical laws are necessary for the truth of the hypothetical sentences, since the latter imply them; but they are not sufficient. All scientific hypothesizing is in the invalid 'third figure' of the implicative syllogism. We neglect here the impossibility, emphasized by Reichenbach and others, of equating even an abstractive sentence or empirical 'law' to a *finite* number of particular observation sentences; this is of importance to philosophers of science but for help in the understanding of theories is of no particular consequence. We shall be assuming the trustworthiness of induction throughout and hence will treat 'direct' observational laws as universal sentences or as sentential functions. One can deduce empirical laws from sentences involving hypothetical constructs, but not conversely. Thus, beginning with the hypothesis that gases are made up of small particles which obey the laws of mechanics, plus certain approximating assumptions about the relation of their sizes to their distances, their perfect elasticity, and their lack of mutual attraction, one can apply mathematical rules and eventually, by direct substitution and equation, lead without arbitrariness to the empirical equation $PV = K$. However, one cannot rigorously reverse the process. That is, one cannot commence with the empirical gas law $PV = K$ and arrive at the full kinetic theory. The mathematics is reversible, granted that certain arbitrary breakups of constants etc., are permitted; but beginning with the empirical law itself there is no basis for these arbitrary breakups. Furthermore, aside from the equations themselves, there are coordinated with these equations certain existence propositions, and assertions about the properties of the entities hypothesized. We state that there exist certain small particles, that they collide with the walls of the container, that the root mean square of their velocities is proportional to the temperature, etc. These assertions can of course not be deduced from the empirical law relating pressure and volume.

This suggests a third distinction between concepts of the two kinds. In the case of abstractive concepts, the quantitative form of the concept, *e.g.*, a measure of its 'amount,' can be derived directly from the empirical laws simply by grouping of terms. In the case of hypothetical concepts, mere grouping of terms is not sufficient. We are less assured of this distinction than of the other two, but we have not been able to think of any exceptions. It seems to us also that, in the case of Hull, this is the point which makes our distinction between hypothetical constructs and intervening variables most obvious. Let us therefore consider Hull's equations as an example.

In *Principles of Behavior*, the influence of certain independent variables such as number of reinforcements, delay in reward, stimulus-response asynchronism, etc., upon response strength is experimentally investigated. In the study of the influence of each of these, the other independent variables are held constant. The experimental findings lead to the formulation of the separate laws of dependence as a set of growth and decay functions. We shall neglect for the moment the complication of drive and of all other variables which intervene between the construct ${}_sH_r$ and the empirical measure of response. That is to say, we shall deal only with the variables introduced in Hull's Postulate 4. The mathematical statement of Postulate 4 is

$${}_sH_r = M\,(1 - e^{-kw})\, e^{-jt}\, e^{-ut'}\, (1 - e^{-iN}). \qquad (5, \text{p. } 178)$$

This equation asserts that habit strength is a certain joint function of four variables which refer to direct empirical quantities—number of reinforcements, delay in reinforcement, amount of reinforcement, and asynchronism between the discriminative stimuli and the response. It is important to see that in this case Hull does not distinguish the four experimentally separated laws combined in the equation by separate concept-names; the only intervening variable introduced is habit strength, which is written as an explicit function of four empirical variables w, t, t', and N. It would be quite possible to introduce an intervening variable referring to, say, the last bracket only; it might be called 'cumulative reinforcement' and it would be a function of only one empirical variable, N. This would be the most reasonable breakdown of habit strength inasmuch as the other three growth functions (two negative) serve merely to modify the asymptote M (5, p. 181). That is to say, given a certain (maintained) rule for the amount of reinforcement given and two time-specifications concerning the constant relation of the response to two other operations, we have determined a parameter m for a dynamic curve describing the course of acquisition of habit strength. The quantity $(1 - e^{-iN})$ (which we are here calling 'cumulative reinforcement') is then an intervening variable which is multiplied by the parameter m in order to determine the value of habit strength after N reinforcements have occurred.

Suppose now that a critic asks us whether our 'cumulative reinforcement' really *exists*. This amounts to asking whether we have formulated a 'correct statement' concerning the relation of this intervening variable to the anchoring (empirical) variables. For since the statement of 'existence' for the intervening variable is so far confined to the equations above, the 'existence' of cumulative reinforcement reduces strictly to the second question. And this second question, as to whether the statement about the intervening variable's relation to the facts is correct, is in turn equivalent to the question, "Are the empirical variables related in such-and-such a way?" In other words, to confirm the equation for habit strength, it is merely necessary to state that (as Hull assumes in his earlier chapters) with drive, etc., constant, some empirical measure of response strength R is a linear function of habit strength. Then we can write directly,

$$R = C\,({}_sH_r) = C\cdot F(w)\, G(t)\, H(t')\, J(N) = Q(w, t, t', N).$$

To confirm or disconfirm this equation is a direct empirical matter. It is possible to multiply out the bracketed quantities in various combinations, so as to make the arbitrary groupings disappear; what will mathematically persist through all such regroupings will be the rather complicated joint function Q of the four empirical variables w, t, t', and N. By various arbitrary groupings and combinations we could define 15 alternative and equivalent sets of intervening variables. Thus, we might multiply out three of the four brackets in the basic equation but for some reason choose to put $e^{-ut'}$ separately into the denominator. This would give us

$$R = \frac{F(w, t, N)}{e^{ut'}}$$

as the particular form for our empirical relation. $F(w, t, N)$ could then be given an appropriate 'intervening variable' name, and the stimulus-response asynchronism t' would then define an intervening variable $e^{ut'}$.

It may be objected that 'habit strength' presumably refers to some state of the organism which is set up by reinforcing N times under specified conditions; whereas $e^{ut'}$ cannot refer to any such state. This seems plausible; but the point is that to establish it as a state, it would be necessary to coordinate to the groupings within equations certain existence propositions, *i.e.*, propositions that do *more* than define a term by saying "Let the quantity $G(x, y, z)$, where x, y, z empirical variables, be designated by the phrase so-and-so." This setting up of existence propositions could presumably be done even for a quantity such as $e^{ut'}$, by referring to such hypothetical processes as, say, diminishing traces in the neural reverberation circuits activated by a certain discriminative stimulus.

In the above example we have considered the fractionation of the intervening variable $_sH_r$ into others. This reasoning can also be extended in the upward direction, *i.e.*, in the direction of fusion rather than fractionation. Let us treat 'habit strength' as Hull would treat our 'cumulative reinforcement,' by not giving it a name at all. It is still possible to set up equations to fit the Perin-Williams data (5, p. 229, 255) without referring to habit strength, writing merely

$$n = F(N, h),$$

where N and h are again both purely empirical variables.

We do not mean to imply that the divisions made by Hull (or Tolman) are of no value. It is convenient to have some term to refer to the result of a certain maintenance schedule, instead of having to say "that part of the general multivariable equation of response strength which contains '*hours since eating to satiety*' as an independent variable." We merely wish to emphasize that in the case of Hull's intervening variables, it is both necessary and sufficient for the truth of his 'theory' about the intervening variables that the empirical facts should be as his equations specify. The latter are merely names attached to certain convenient groupings of terms in his empirically fitted equations. It is always possible to coordinate to these quantities, which as written mathematically contain parameters and experimental variables only, certain existence propositions which would automatically make the construct 'hypothetical' rather than 'abstractive'. This giving of what Reichenbach calls 'surplus meaning' automatically destroys the equivalence between the empirical laws and the theoretical construct. When habit strength *means* the product of the four functions of w, t, t', and N, then if the response strength is related to these empirical variables in the way described, habit strength 'exists' in the trivial sense that the law holds. Our confidence in the 'correctness' of the intervening variable formulation is precisely as great as our confidence in the laws. When, however, habit strength means not merely this product of empirical functions but something more of a neural or other physiological nature, then the theory could be false even if the empirical relations hold.

It seems to us that Tolman himself, in using one of Hull's serial conditioning diagrams as a set of intervening variables (11, p. 13), departs from his original definition. He has first described the situation in which the 'behavior ratio' is a complex function f_1 of the independent experimental variables. He goes on to say,

"A theory, as I shall conceive it, is a set of intervening variables. These to-be-inserted intervening variables are 'constructs' which we, the theorists, evolve as a useful way of breaking down into more manageable form the original complete f_1 function" (11, p. 9).

His reason for introducing intervening variables does not seem to us very cogent as he states it. He says that empirically establishing the form of f_1 to cover the effects on behavior of all the permutations and combinations of the independent variables would be a 'humanly endless task.' If this means that all of the verifying instances of a continuous mathematical function cannot be empirically achieved it is true; but that is equally true for a function of one variable only. In order to utilize the proposed relationship between Tolman's function f_3 (11, p. 10) which describes the relation of the behavior to the intervening variables, it is still necessary to establish empirically that the relationship holds—which amounts essentially to trying several of the infinitely many permutations and combinations (as in the Perin-Williams study) until we are inductively satisfied by ordinary scientific standards.

However cogent the arguments for intervening variables may be, it seems clear from Tolman's description that they are what we are calling *abstractive* rather than hypothetical. His notion of them involves nothing which is not in the empirical laws that support them. (We may speak of 'laws' here in the plural in spite of there being just the single function f_1, just as Boyle's and Charles' laws are distinguished in addition to the more general gas law $PV/T = R$.) For Tolman, the merit of an intervening variable is of a purely 'summarizing' character. One can determine the function f_1 by parts, so to speak (11, p. 17), so that the effect of a given maintenance schedule upon one part of f_1 may be referred to conveniently as *drive*. For a given drive, we can expect such-and-such behavior ratios in a diversity of situations defined by various combinations of the other independent variables.

It has been observed earlier that in introducing one of Hull's well-known serial conditioning diagrams as an example of intervening variables outside Tolman's own system, we see a departure from the definition Tolman gives. The Hull diagrams contain symbols such as r_g (fractional anticipatory goal response) and s_g (the proprioceptive impulses produced by the movements constituting r_g). These symbols refer to hypothetical processes within the organism, having an allegedly real although undetermined neuromuscular locus. These events are in principle directly observable. In fact, here the case for speaking of an objective reality is even stronger than Reichenbach's examples of electrons, molecules, etc.; since even the criterion of *technical* verifiability, admitted by all positivists to be too strong a restriction, would not exclude these hypotheses as empirically meaningless. Even without penetrating the organism's skin we have some direct observational evidence of r_g in the work of Miller (7). Whether r_g occurs and actually plays the role described is not relevant here; the point is that the diagrams and verbal explanations of Hull involve the supposition that it does. He assumes the existence of certain processes which are not logically implied by the empirical laws in the sense of strict equivalence. Even if, by using the notion of fractional anticipatory goal response, Hull deduced all of the empirical laws relating independent and dependent variables, alternative hypotheses could be offered. Because of the 'surplus meaning' contained in concepts like r_g and s_g, these concepts are not really 'anchored' to the facts in the sense implied by Tolman's definition of intervening variables or by Hull's diagram on page 22 of the *Principles*. Hull states in reference to this diagram,

> "When an intervening variable is thus securely anchored to observables on both sides it can be safely employed in scientific theory" (5, p. 22)

We presume that Hull means in this statement that the anchoring in question is not only a sufficient but a necessary condition for scientific admissibility. We feel that the criterion is too strong, assuming that the structure of modern physical science is to be allowed. This sort of anchoring makes the intervening variable strictly reducible to the empirical laws, which is, to be sure, what Tolman's original definition implied. But it excludes such extremely fruitful hypotheses as Hull's own fractional anticipatory goal responses, for which the strict reducibility does not exist.

It occurs to us also in this connection that Hull seems to have moved in the direction of Skinner and Tolman in his treatment of intervening variables. The use of the postulate-theorem approach is maintained more as a form in the *Principles* than as an actual instrument of discovery. In this respect, the *Principles* is much less like Hull's Newtonian model than was the *Mathematico-deductive theory of rote learning*. The justification of 'postulates' in the usual sense is their ability to mediate deductions of empirical laws which are then verified. In the *Principles*, the 'postulates' are verified directly, by the experimental device of holding all variables constant except the one for which we want to find a law. This is quite unlike the derivation of the gas law in physics. The only sense in which any postulates are 'assumed' is in the assumption, referred to by Hull on page 181 of the *Principles*, that the separately verified parts of Postulate 4 will in fact operate according to his equation 16 when combined. This is certainly a 'postulate' only in a very attenuated sense, since it amounts essentially to an empirical extrapolation which can be verified directly, as Hull suggests.

At this point any distinction between the type of theory advocated by Hull and that advocated by Skinner or Tolman would seem to disappear, except for the relatively non-contributory 'neural' references contained in the verbal statement of Hull's postulates. Insofar as this neural reference is taken seriously, however, we are still dealing with concepts of a hypothetical rather than abstractive character. There are various places in Hull's *Principles* where the verbal accompaniment of a concept, which in its mathematical form is an intervening variable in the strict (Tolman) sense, makes it a hypothetical construct. Thus, the operational definition of a *pav* of inhibition (5, p. 281) would seem merely to mean that when we know from the independent variables that the combined habit strength and drive, together with a discriminative stimulus located so many j.n.d.'s from the original, would yield a reaction potential of so many wats, it requires an equal number of pavs of inhibition to yield an effective reaction potential of zero. However, in the accompanying verbal discussion (5, p. 281) Hull refers to the removal of the inhibitory substance by the blood stream passing through effector organs as determining the quantitative law of spontaneous loss of inhibition as a function of time. 'Afferent neural interaction' is another example of a concept which is mathematically represented as a relation of intervening variables in Tolman's sense, but to which are coordinated verbal statements that convey the surplus meaning and make it an hypothesis.

The question might be raised, whether this is not always the difference—that the mathematical assertions are definitive of intervening variables but the verbal additions lend the hypothetical character to such concepts. We do not believe this is the essential difference. There are mathematical expressions whose meaning is not defined in the absence of verbal existential accompaniment, because the quantities involved refer to non-observational (*i.e.*, hypothetical) processes or entities. There are other mathematical expressions for which this is not true, since their component symbols have direct observational reference. In the case of our 'cumulative reinforcement' term $(1 - e^{-iN})$, no coordinated existential proposition is required. We simply say, "Response probability is such-and-such a multivariate function of such-and-such experimental variables. Within this function can be isolated a simple growth function of one variable, whose value as a function of N is referred to as *cumulative reinforcement*." This may be taken as an adequate reference for $(1 - e^{-iN})$. On the other hand, in the derivation of the law $PV = K$ there occur statements such as "When the gas is maintained at the same temperature, $mv^2 / 2$ does not change." Neither m nor v is an empirical variable. This statement does not tell us anything *until* we are informed that v refers to the velocity which each molecule of the gas could be assumed to have in order that their mean kinetic energy should be what it is. In other words, in the derivation of the gas laws from kinetic theory there occur mathematical assertions whose meaning is unclear without the accompanying existence assertions, and *which cannot be utilized to take the subsequent mathematical steps in the chain of inferences unless these assertions are included*. Thus, to get from a purely mathematical statement that a molecule on impact conserves all of its momentum, to a mathematical statement whose terms refer to the empirical concept of 'pressure on the walls,' it is necessary to know (from the accompanying verbal description) that in the equations of derivation, m refers to the mass of a hypothetical particle that strikes the wall, v to its velocity, and so on. This example shows that some mathematical formulations are themselves incomplete in the sense that they cannot mediate the desired deductions unless certain existential propositions are stated alongside, so as to render certain necessary substitutions and equations legitimate. Therefore it is not merely the matter of mathematical form that distinguishes a 'pure' intervening variable from a hypothesis.

In the second place, it seems to us that the use of verbal statements without mathematical formulations does not guarantee that we are dealing with a hypothetical construct rather than an intervening variable. Consider Skinner's definition of emotion as a 'state of the organism' which alters the proportionality between reserve and strength. This is not defined as a direct proportionality, and in fact Skinner nowhere deals with its quantitative form. No mathematical statement is given by him; yet we would contend that the use of the word 'state' does not in any way make the notion of emotion existential, any more than drive is existential in Skinner's usage. The 'state' of emotion is not to be described in any way except by specifying (a) The class of stimuli which are able to produce it and (b) The effects upon response strength. Hence emotion for Skinner is a true intervening variable, in Tolman's original sense. We conclude from these examples that whether a given concept is abstractive or hypothetical is not merely a matter of whether it is an equation with or without accompanying verbal exposition.

On the basis of these considerations, we are inclined to propose a linguistic convention for psychological theorists which we feel will help to clarify discussion of these matters. We suggest that the phrase 'intervening variable' be restricted to the original use implied by Tolman's definition. Such a variable will then be simply a quantity obtained by a specified manipulation of the values of empirical variables; it will involve no hypothesis as to the existence of nonobserved entities or the occurrence of unobserved processes; it will contain, in its complete statement for all purposes of theory and prediction, no words which are not definable either explicitly or by reduction sentences in terms of the empirical variables; and the validity of empirical laws involving only observables will constitute both the necessary and sufficient conditions for the validity of the laws involving these intervening variables. Legitimate instances of such 'pure' intervening variables are Skinner's *reserve*, Tolman's *demand*, Hull's *habit strength*, and Lewin's *valence*. These constructs are the behavioral analogue of Carnap's 'dispositional concepts' such as solubility, resistance, inflammability, etc. It must be emphasized that the setting up of a definition or reduction for an intervening variable is not a wholly arbitrary and conventional matter. As Carnap has pointed out, it often happens that we give alternative sets of reduction sentences for the same dispositional concept; in these cases there is empirical content in our statement even though it has a form that suggests arbitrariness. The reason for this is that these separate reductions for a given dispositional concept imply that the empirical events are themselves related in a certain way. The notion of amount of electric current can be introduced by several different observations, such as deposition of silver, deflection of a needle, hydrogen separated out of water, and so on. Such a set of reductions has empirical content because the empirical statements together with the reductions must not lead to contradictions. It is a contingent fact, not derivable from definitions alone, that the deposition of silver will give the same answer for 'amount of current' as will the deflection of a needle. A similar problem exists in Hull, when he sets up 'momentary effective reaction potential' as the last intervening variable in his chain. In the case of striated muscle reactions, it is stated that latency, resistance to extinction, and probability of occurrence of a response are all functions of reaction potential. Neglecting behavior oscillation, which does not occur in the formulation for the second two because they involve many repetitions of the situation, this means that the empirical variables must be perfectly correlated (non-linearly, of course). The only possible source of variation which could attenuate a perfect correlation between probability of occurrence and resistance to extinction would be actual errors of experimental measurement, since there are no sources of uncontrolled variation left within the organism. If we consider average latency instead of momentary latency (which is a function of momentary effective reaction potential and hence varies with behavioral oscillation), latency and resistance to extinction should also be perfectly correlated. It remains to be seen whether the fact will support Hull in giving simultaneously several reductions for the notion of reaction potential.

As a second linguistic convention, we

propose that the term 'hypothetical construct' be used to designate theoretical concepts which do *not* meet the requirements for intervening variables in the strict sense. That is to say, these constructs involve terms which are not wholly reducible to empirical terms; they refer to processes or entities that are not directly observed (although they need not be in principle unobservable); the mathematical expression of them cannot be formed simply by a suitable grouping of terms in a direct empirical equation; and the truth of the empirical laws involved is a necessary but not a sufficient condition for the truth of these conceptions. Examples of such constructs are Guthrie's M.P.S.'s, Hull's r_g's, S_d's, and *afferent neural interaction*, Allport's *biophysical traits*, Murray's *regnancies*, the notion of 'anxiety' as used by Mowrer, Miller, and Dollard and others of the Yale-derived group, and most theoretical constructs in psychoanalytic theory. Skinner and Tolman seem to be almost wholly free of hypothetical constructs, although when Skinner invokes such notions as the 'strain on the reserve' (10, p. 289) it is difficult to be sure.

We do not wish to seem to legislate usage, so that if the broader use of 'intervening variable' has become stuck in psychological discourse, we would propose alternatively a distinction between intervening variables of the 'abstractive' and of the 'hypothetical' kind. Since our personal preference is for restricting the phrase *intervening variables* to the pure type described by Tolman, we shall follow this convention in the remainder of the present paper.

The validity of intervening variables as we define them cannot be called into question except by an actual denial of the empirical facts. If, for example, Hull's proposed 'grand investigation' of the Perin–Williams type should be carried out and the complex hyperspatial surface fitted adequately over a wide range of values (5, p. 181), it would be meaningless to reject the concept of 'habit strength' and still admit the empirical findings. For this reason, the only consideration which can be raised with respect to a given proposed intervening variable, when an initial defining or reduction equation is being written for it, is the question of convenience.

In the case of hypothetical constructs, this is not so clear. Science is pursued for many reasons, not the least of which is *n Cognizance*. Since hypothetical constructs assert the existence of entities and the occurrence of events not reducible to the observable, it would seem to some of us that it is the business of a hypothetical construct to be 'true'. It is possible to advance scientific knowledge by taking a completely 'as if' attitude toward such matters, but there are always those whose theoretical-cognitive need dictates that existential propositions should correspond to what is in fact the case. Contemporary philosophy of science, even as represented by those who have traditionally been most cautious about discussing 'truth' and most highly motivated to reduce it to the experiential, gives psychologists no right to be dogmatic about the 'as if' interpretation of theoretical knowledge (*cf.* especially Carnap, 4, p. 598, Kaufmann, 6, p. 35, Russell, 9, Introduction and Chapter XXI, and Reichenbach, 8, *passim*). We would find it rather difficult to defend the ingenious conditioning hypotheses developed in Hull's series of brilliant papers (1929-) in the PSYCHOLOGICAL REVIEW on the ground that they merely provide a "convenient shorthand summarization of the facts" or are of value in the 'practical manipulation' of the rat's behavior. We suspect that Professor Hull himself was motivated to write these articles because he considered that the hypothetical events represented in his diagrams may have actually occurred and that the occurrence of these events represents the underlying truth about the learning phenomena he dealt with. In terms of practical application, much (if not most) of theoretical psychology is of little value. If we exclude the interesting anecdotes of Guthrie, contemporary learning theory is not of much use to school teachers. As a theoretical enterprise, it may fairly be demanded of a theory of learning that those elements which are 'hypothetical' in the present sense have some probability of being in correspondence with the actual events underlying the behavior phenomena, *i.e.*, that the assertions about hypothetical constructs be true.[1]

Another consideration may be introduced here from the standpoint of future developments in scientific integration. Even those of us who advocate the pursuit of behavioral knowledge on its own level and for its own sake must recognize that some day the 'pyramid of the sciences' will presumably catch up with us. For Skinner, this is of no consequence, since his consistent use of intervening variables in the strict sense genuinely frees him from neurophysiology and in fact makes it possible for him to impose certain conditions upon neurophysiological explanations (10, pp. 429-431). Since he hypothesizes nothing about the character of the inner events, no finding about the inner events could prove disturbing to him. At most, he would be able to say that a given discovery of internal processes must not be complete because it cannot come to terms with his (empirical) laws. But for those theorists who do not confine themselves to intervening variables in the strict sense, neurology will some day become relevant. For this reason it is perhaps legitimate, even now, to require of a hypothetical construct that it should not be manifestly unreal in the sense that it assumes inner events that cannot conceivably occur. The 'as if' kinds of argument sometimes heard from more sophisticated exponents of psychoanalytic views often seem to ignore this consideration. A concept like *libido* or *censor* or *super-ego* may be introduced initially as though it is to be an intervening variable; or even less, it is treated as a merely conventional designation for a class of observable properties or occurrences. But somewhere in the course of theoretical discussion, we find that these words are being used as hypothetical constructs instead. We find that the libido has acquired certain hydraulic properties, or as in Freud's former view, that the 'energy' of libido has been converted into 'anxiety.' What began as a name for an intervening variable is finally a name for a 'something' which has a host of causal properties. These properties are not made explicit initially, but it is clear that the concept is to be used in an explanatory way which requires that the properties exist. Thus, libido may be introduced by an innocuous definition in terms of the 'set of sexual needs' or a 'general term for basic strivings.' But subsequently we find that certain puzzling phenomena are *deduced* ('explained') by means of the various prop-

[1] It is perhaps unnecessary to add that in adopting this position we do not mean to defend any form of metaphysical realist thesis. The ultimate 'reality' of the world in general is not the issue here; the point is merely that the reality of hypothetical constructs like the atom, from the standpoint of their logical relation to grounds, is not essentially different from that attributed to stones, chairs, other people, and the like. When we say that hypothetical constructs involve the notion of 'objective existence' of actual processes and entities within the organism, we mean the same sort of objective existence, defined by the same ordinary criteria, that is meant when we talk about the objective existence of Singapore. The present discussion operates within the common framework of empirical science and common sense and is intended to be metaphysically neutral.

erties of libido, *e.g.*, that it flows, is dammed up, is converted into something else, tends to regress to earlier channels, adheres to things, makes its 'energy' available to the ego, and so on. It is naive to object to such formulations simply on the ground that they refer to unobservables, or are 'hypothetical,' or are not 'statistical.' None of these objections is a crucial one for any scientific construct, and if such criteria were applied a large and useful amount of modern science would have to be abandoned. The fundamental difficulty with such theories is two-fold. First, as has been implied by our remarks, there is the failure explicitly to announce the postulates concerning existential properties, so that these are introduced more or less surreptitiously and *ad hoc* as occasion demands. Secondly, by this device there is subtly achieved a transition from admissible intervening variables to inadmissible hypothetical constructs. These hypothetical constructs, unlike intervening variables, are inadmissible because they require the existence of entities and the occurrence of processes which cannot be seriously believed because of other knowledge.

In the case of libido, for instance, we may use such a term legitimately as a generic name for a class of empirical events or properties, or as an intervening variable. But the allied sciences of anatomy and physiology impose restrictions upon our use of it as a hypothetical construct. Even admitting the immature state of neurophysiology in terms of its relation to complex behavior, it must be clear that the central nervous-system does not in fact contain pipes or tubes with fluid in them, and there are no known properties of nervous tissue to which the hydraulic properties of libido could correspond. Hence, this part of a theory about 'inner events' is likely to remain metaphorical. For a genuine intervening variable, there is no metaphor because all is merely shorthand summarization. For hypothetical constructs, there is a surplus meaning that is existential. We would argue that dynamic explanations utilizing hypothetical constructs ought not to be of such a character that they *have* to remain only metaphors.

Of course, this judgment in itself involves a 'best guess' about the future. A hypothetical construct which seems inherently metaphorical may involve a set of properties to which hitherto undiscovered characteristics of the nervous system correspond. So long as the propositions about the construct are not stated in the *terms* of the next lower discipline, it is always a possibility that the purely formal or relational content of the construct will find an isomorphism in such characteristics. For scientific theories this is enough, since here, as in physics, the associated mechanical imagery of the theorist is irrelevant. The tentative rejection of libido would then be based upon the belief that no neural process is likely to have the *combination* of formal properties required. Strictly speaking, this is always problematic when the basic science is incomplete.[2]

SUMMARY

1. At present the phrases 'intervening variable' and 'hypothetical construct' are often used interchangeably, and theoretical discourse often fails to distinguish what we believe are two rather different notions. We suggest that a failure to separate these leads to fundamental confusions. The distinction is between constructs which merely abstract the empirical relationships (Tolman's original intervening variables) and those constructs which are 'hypothetical' (*i.e.*, involve the supposition of

[2] We are indebted to Dr. Herbert Feigl for a clarification of this point.

entities or processes not among the observed).

2. Concepts of the first sort seem to be identifiable by three characteristics. First, the statement of such a concept does not contain any words which are not reducible to the empirical laws. Second, the validity of the empirical laws is both necessary and sufficient for the 'correctness' of the statements about the concept. Third, the quantitative expression of the concept can be obtained without mediate inference by suitable groupings of terms in the quantitative empirical laws.

3. Concepts of the second sort do not fulfill any of these three conditions. Their formulation involves words not wholly reducible to the words in the empirical laws; the validity of the empirical laws is not a sufficient condition for the truth of the concept, inasmuch as it contains surplus meaning; and the quantitative form of the concept is not obtainable simply by grouping empirical terms and functions.

4. We propose a linguistic convention in the interest of clarity: that the phrase *intervening variable* be restricted to concepts of the first kind, in harmony with Tolman's original definition; and that the phrase *hypothetical construct* be used for those of the second kind.

5. It is suggested that the only rule for proper intervening variables is that of convenience, since they have no factual content surplus to the empirical functions they serve to summarize.

6. In the case of hypothetical constructs, they have a cognitive, factual reference in addition to the empirical data which constitute their support. Hence, they ought to be held to a more stringent requirement in so far as our interests are theoretical. Their actual existence should be compatible with general knowledge and particularly with whatever relevant knowledge exists at the next lower level in the explanatory hierarchy.

RERERENCES

1. Benjamin, A. C. *An introduction to the philosophy of science*. New York: Macmillan, 1937.
2. Carnap, R. Testability and meaning, Parts I-III. *Phil. Sci*., 1936, **3**, 419-471.
3. Carnap, R. Testability and meaning, Part IV. *Phil. Sci*., 1937, **4**, 1-40.
4. ——. Remarks on induction and truth. *Phil. & phenomenol. res*., 1946, **6**, 590-602.
5. Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
6. Kaufmann, F. *Methodology in the social sciences*. London: Oxford University Press, 1944.
7. Miller, N. E. A reply to 'Sign-Gestalt or conditioned reflex.' PSYCHOL. REV., 1935, **42**, 280-292.
8. Reichenbach, H. *Experience and prediction*. Chicago: University of Chicago Press, 1938.
9. Russell, B. *Inquiry into meaning and truth*. New York: Norton, 1940.
10. Skinner, B. F. *Behavior of organisms*. New York: Appleton-Century, 1938.
11. Tolman, E. C. The determiners of behavior at a choice point. PSYCHOL. REV., 1938, 45, 1-41.

pdf by ljy August 2002