

MacCorquodale, K., & Meehl, P. E. (1954). E. C. Tolman.  
In W. K. Estes, S. Koch, K. MacCorquodale, P. E. Meehl, C. G.  
Mueller, W. N. Schoenfeld, & W. S Verplanck, *Modern learning  
theory* (pp. 177-266). New York: Appleton-Century-Crofts.

#031

## *Edward C. Tolman*



**KENNETH MacCORQUODALE  
and PAUL E. MEEHL**

I. INTRODUCTION .....	1
II. STRUCTURE OF TOLMAN'S THEORY .....	3
III. ABILITY OF TOLMAN'S THEORY TO ACCOUNT FOR FACTS NOT PREDICTABLE FROM COMPETING THEORIES .....	31
IV. SOME TENTATIVE PROPOSALS AS TO FORMALIZATION .....	51
V. EXCURSUS: THE RESPONSE CONCEPT .....	57
VI. APOLOGY TO TOLMANITES AND OTHERS .....	99
SUMMARY AND CONCLUSIONS .....	113
Bibliography .....	114

### **I. INTRODUCTION**

THE PRESENT DISCUSSION of Tolman's theory is based upon his statements of it as they have appeared beginning with 1932 when *Purposive Behavior in Animals and Men* (128) was published; the most recent statement we have considered is the 1949 article: "There is more than one kind of learning" (151). We have not included in our bibliography or in our discussion Tolman's more recent views on the conceptualization of social behavior; neither have we considered the several constructive commentaries and interpretations authored by others, such as White (185). Our reasons for not treating Tolman's most recent writings stem in part from our understanding that he is currently engaged, with Postman, in a major restatement of his views. We are speaking in this section, then, of Tolman as he has been under-

stood for the past twenty years. Whether our criticisms will still be up-to-date after publication of his new book depends upon whether he proposes a conceptually new statement or a restatement of the present concepts with greater felicity of expression, and remains to be seen.

Our aim has been two-fold: first, to display and critically examine Tolman's system as it is; secondly, to state a set of more formal "laws" which introduce those concepts we consider to be most central to Tolman's views and which are intended to preserve the flavor while increasing the formalization of an expectancy theory.

The system will be considered according, generally, to the outline described in Chapter I.\* Considerable emphasis will be given to the aspects of Tolman's theory which differentiate it from the other theories considered in this volume, and for this reason it is important to stress that much of the content of his 1920 to 1932 publications, and even of *Purposive Behavior in Animals and Men*, is in substantial agreement with the views of *S-R* and *S-R* reinforcement theorists. To mention one obvious example, most psychologists trained since the early thirties, whether Tolmanites or not, find it so natural to think in terms of "reinforcements," "rewards" or "goal-objects" when speaking of instrumental learning, that we easily forget the importance of Tolman's work in reducing the plausibility of the variant of pure frequency-recency connectionism which was held by many prior to around 1930. There are some characterizations of Tolman's theory, often said to differentiate him from other theorists, which will not be emphasized here at all. An example is the frequent reference to Tolman as a Gestalt or field theorist. The justification for our omission of this emphasis will be made in a later part of the discussion.

In several places we will point out a difficulty and include a comment to the effect that "this problem is not peculiar to Tolman." Such indications of common difficulties, which appear especially in

---

\* [Chapter 1 of *Modern Learning Theory*, 1954. The book resulted from a seminar held at Dartmouth College in 1950, referred to later in this chapter as the "Dartmouth Conference."—LJY]

Part IV, are not to be interpreted as mere *tu quoque* or as denials of problems. Needless to say, we have not pretended to suggest directions of solution for all problems; and since our subject is Tolman, we have devoted particular attention to those questions which arise from the special orientation that distinguishes him as a systematist. This seemed to us the more necessary because Tolman is sometimes criticized for lacks or ambiguities which *seem* to be peculiar to him, but on careful examination are found to be present, although obscured, in competing systems.

## II. THE STRUCTURE OF TOLMAN'S THEORY

### *Data Language*

Considering first Tolman's use of a data language, the first chapter of *Purposive Behavior in Animals and Men* contains statements which appear to attempt to make such words as "purpose" and "cognition" a part of the data language. For example: "Behavior qua behavior has descriptive properties of its own ... getting-to or getting-from" (p. 10); "These purposes and cognitions are of its immediate descriptive warp and woof ... behavior reeks of purpose and of cognition" (p. 12). However, these may be understood as devices used by a theorist in an introductory chapter of a theoretical work, proposing a behaviorism in which such apparently non-behavioral words as "purpose" and "cognition" will appear with behavioral definitions. That he intends to define such words as "purpose" in terms of certain features of behavior seems clear.

When actual experiments are being discussed, a relatively clean, neutral data language seems to be maintained (although when theoretical derivations are being performed, there are some confusions, as will be seen below). As an example of an unequivocal description of experimental events, the following appears in *Purposive Behavior in Animals and Men* (p. 48):

In this maze he [Blodgett] ran three groups of hungry rats. Each group had one trial per day. The *control group* (Group I) was allowed to eat in the usual manner for three minutes in the food-box at the end of the maze. They were then immediately fed the remainder of their day's ration in another cage (not the home cage). The *first experimental group* (Group II) ran the maze for the first six days, without immediate reward. That is, at the end of the maze they were confined in the exit-box without food for two minutes and obtained their day's ration only one hour, or more, afterwards in another cage (not the home cage). After six days of such running, this group, on the seventh day, suddenly found food in the exit-box, and continued so to find it on all subsequent days. A *second experimental group* (Group III) ran the maze without food at the exit-box for two days. For these two days they, like Group II, obtained their day's ration only one hour or more afterwards in another cage. On the third day, however, food was given them in the exit-box and they continued to find it there on all subsequent days.

In the above, the word "hungry," though not further defined here, has become a frequently used and accepted ellipsis; the word "suddenly" is perhaps extravagant, but seems to carry no illicit descriptive connotations here.

In general, Tolman's departures from explicitness in the data language are those which can either be easily filled in (e.g., "hungry") or are permissible as being defined in another discipline (e.g., "sunflower seeds" from botany, or "gravitational" from physics).

### ***Operational Definitions***

Tolman's use of operational reduction is seen most explicitly in the article "Studies in Spatial Learning I. Orientation and the short-cut" (173). The following is the definition of an expectancy "revised to better express the original intent of the senior author" (p. 15):

When we assert that a rat expects food at  $L$ , what we assert is that *if* (1) he is deprived of food, (2) he has been trained on path  $P$ , (3) he is now put on path  $P$ , (4) path  $P$  is now blocked, and (5) there are other paths which lead away from path  $P$ , one of which points directly to location  $L$ ,

*then* he will run down the path which points directly to location *L*.

When we assert that he does *not* expect food at location *L*, what we assert is that, under the same conditions, he will *not* run down the path which points directly to location *L*.

Restated as a conditioned definition, this becomes:

*If* *x* is deprived of food and *x* has been trained on path *P* and *x* is now put on path *P* and path *P* is now blocked and there are other paths which lead away from path *P*, one of which points directly to location *L*, *then* (*x* runs down the path which points directly to location *L*  $\equiv$  *x* expected food at location *L*).

Three comments on this formulation may be offered. First, as a reduction sentence it is incomplete in several respects. It fails to specify, for example, what else may have been in the rat's history, and what restrictions are placed on path *P*'s stimulus properties in contrast to other paths. In the case, not excluded by this definition, in which path *P* is white and it is blocked, and there is among the "other" paths one white path (among non-white paths) which leads to location *L*, *then* if the rat takes "the path which leads to location *L*" are we prepared to assert that he "expects food at *L*"?

Secondly, the significance of the negative case is somewhat puzzling by this definition; it would seem, as stated in the singular, that the rat which *fails* to turn down path *P* does *not* expect food there. Does an expectancy theorist really mean to assert this? Along the same lines, the statement of the operational reduction in terms of the behavior of the individual rat suggests a disturbing statistical consideration: If 12 out of 100 rats run down path *P* from among 8 paths, shall we assume that these 12 had expectancies and the remainder did not? Now, it might be objected that these criticisms are petty, that *of course* it is intended that the paths should not be different in other than spatial respects, and *of course* ordinary statistical significance is demanded. Since, however, the intent in the use of operational definitions is to make explicit the complete set of relevant characteristics of an event, the above definitions must be considered incomplete.

Our third comment refers less to the adequacy of this definition as an example of the operational method, than to the rationale of its appearance in a theoretical structure such as Tolman's. He says: "... I do not hold, as do most behaviorists, that all learning is, as such, the attachment of responses to stimuli. Cathexes, equivalence beliefs, field expectancies, field-cognition modes and drive discriminations are not, as I define them, stimulus-response connections. They are central phenomena, each of which may be expressed by a variety of responses" (151, p. 146). By giving a directly operational definition of a central state, such as an expectancy, in terms of *a* locomotion, Tolman is in danger of losing precisely that advantage which a centralism has over a peripheralism. In Hull's 1943 *Principles of Behavior* (38, p. 383) the basic learning construct,  $sH_R$ , is four steps removed in the intervening variable chain from his "operational" measures of actual momentary response strength. Speaking loosely for the moment, one might suppose that one of the advantages of a more centralist approach is in the matter of some such *causal distance* of the hypothetical cognitive unit from the effector-activity. The kind of means-to-ends appropriateness shown by behavior occurring in altered environmental circumstances which has traditionally intrigued the expectancy theorist is, from a methodological viewpoint, an argument for increasing this causal distance, and for getting, so to speak, more "play" into the system. See, for instance, Tolman's quotation from Gilhousen on the character of crayfish learning (128, p. 18). The question of an alternative mode of definition of such central states will be considered later in the discussion of Tolman's principal constructs and their appearance in derivation of experimental results.

### ***Dependent and Independent Variables***

In his presidential address to the American Psychological Association in 1937 (141), Tolman has formally listed the dependent and independent variables with which he proposes to deal. The following are the independent variables listed with briefer definitions of the

symbols:

$\Sigma OBO$  : The sum of the preceding occasions on which stimulus complex  $O$ , has been followed by behavior  $B_x$ , which has been followed by stimulus complex  $O_y$ .

$M$  : Maintenance schedule.

$G$  : Appropriateness of the goal object.

$S$  : Types and modes of stimuli provided.

$R$  : Types of motor response required.

$P$  : Pattern of preceding and succeeding maze units.

There is also a class of Individual Differences variables on the independent variable side:

$H$  : Heredity.

$A$  : Age.

$T$  : Training.

$E$  : Endocrine, drug and vitamin conditions.

There are several noteworthy characteristics of the above “independent” variables.  $G$ , the appropriateness of the goal-object, depends upon an adequate definition of “appropriateness” if it is to be firmly anchored in the data language. Although it is not so anchored by Tolman, it appears as a general term for such characteristics of an incentive as its relation to the drive state (or maintenance schedule), its hardness, bigness, saltiness, etc.; in brief, to characteristics which are expressible in a data language. May we not say that Hull would presumably encounter a similar difficulty in defining his “(amount and) quality of the reinforcing agent” variable, and left the *quality* dimension largely undefined?

More striking is the fact that several variables which are ordinarily characterized as being intervening or dependent have, in this list, found their way to the independent variable side of the causal system. Thus “ $R$ , the types of motor response required,” which refers to whatever class of topographies the experimenter rewards, is an independent variable only in this sense: that the experimenter can manipulate it by progressively altering its frequency of occurrence through the learning

itself. This variable must occur, or at least recur, as a dependent variable. Moreover, it is difficult to discriminate among the three variables referring to the effects of past training, which, as *state* variables are doubtfully appropriate for this list. These are  $\Sigma OBO$ ,  $T$  and  $P$ . Insofar as  $P$ , the pattern of preceding and succeeding maze units can affect behavior at a choice point only in consequence of previous  $OBO$ 's in those units and in the given maze, while  $T$  refers to previous  $OBO$ 's in other mazes and might be referred to as transfer effects, these three variables are not discrete.

For the choice point example, Tolman specifies the *behavior ratio* as the dependent variable. The behavior ratio is defined as

$$\frac{\text{frequency of turn into alley } L}{\text{total frequency of turns into alleys } L \text{ \& } R}$$

It will be noted that in practice, it is difficult or impossible actually to compute the behavior ratio for a rat at a choice point, since one rat cannot, during the course of learning, be made to choose repeatedly at the same choice point at the same response strength. This is one of the several defects of the maze as an instrument, as Skinner has pointed out. For this reason, the literal definition of the behavior ratio is rarely applied, and for it is substituted a count of the number of entries into all culs by a rat (his "score"), when there are several culs involved. When the "behavior ratio" needs to be determined for a specified choice point, determinations of left and right choices for *groups* of rats may be made, but this gives no direct indication of the strengths for individual organisms. The mathematical relationship of the choice ratio to the strengths of the two competing responses is not indicated.

### ***Relation Between Empirical Area Covered and Orientative Attitudes***

The dependent and independent variables listed above are specifically appropriate for the discussion of the maze running case, and this is the empirical area to which Tolman has largely restricted himself. It should be pointed out, however, that in Part V of *Purposive Behavior*

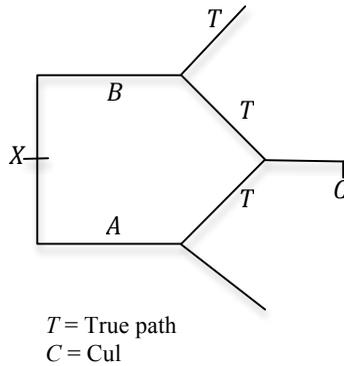
*in Animals and Men* Tolman has systematically considered conditioned reflex and trial-and-error learning (as these were understood in 1932), and tried to show that his vocabulary can be applied to them. He has recently discussed psychoanalytic mechanisms such as fixation, regression and displacement (150) and social behavior (153). In neither case, however, has he specifically derived social or psychoanalytic phenomena as formal consequences of his theory.

### ***Hypothetical Constructs***

Tolman's definitions of the kinds of constructions which he intends to use indicate that he wished (1932) to avoid hypothetical constructs in the sense proposed by MacCorquodale and Meehl (56), in which a construct acquires attributes which exceed its original defining operations.\* He proposed instead (134, 135, 141) to define his intervening variables by an inductive breaking down of the function  $f_1$  (relating the dependent and independent variables) into intervening variables  $a, b, c$ , with a set of  $f_2$  functions to connect these with the independent variables on the one hand and a set of  $f_3$  functions to connect them with the dependent variables on the other. Tolman has felt it important to emphasize that his usage of such intervening terms as "expectancy" and "cognition" is to be understood in terms of the establishing operations, since these words have such extensive lay connotations; however, as will be pointed out below, the power of these concepts for Tolman in the derivation of experimental results is severely lessened by their exclusively "intervening variable" character. Some of the empirical support Tolman adduces for his constructs appears to be support only because the common-sense kind of surplus meaning is informally attached to these theoretically "neutralized" terms.

---

\* He has, however, recently (152) expressed a preference for hypothetical constructs as defined by us.

FIG. 2<sup>†</sup>

To test the actual reducibility of these construct-terms to the physical or data language two approaches will be taken here: first, one can find *concrete* instances of the use of a term and test the adequacy of its reduction in the context of its occurrence; second, it is interesting to study how concept words as defined in the glossary of *Purposive Behavior in Animals and Men* are *there reduced* to the data language.

As an example of the first kind, refer to *Purposive Behavior in Animals and Men*, Chapter XI: "Inference—the means-end-field," p. 175. In an experiment by Buel, two groups of rats ran the same maze made up of several similar units (Fig. 2) with the difference that for one group, barrier *X* was present so that the culs *A* and *B* constituted separate blinds; for the other group, the barrier *X* was removed so that culs *A* and *B* were continuous, or "constituted the reverse ends of one and the same detour." The rats of the second group made fewer entrances into *B* than did the rats of the first group. "In other words, whereas *B* has to be learned as a separate independent blind for the animals of Group I, for Group II it can be *recognized as but the other end of A, already entered.*" If we are interested in the definitions of the terms in the italicized part of the sentence (not

<sup>†</sup> [This is the first figure for this chapter of the book.]

italicized in the original) we discover that *recognize* is not defined here, nor does it appear in the glossary or index; and surely it is not a word in the data language. *The other end of A* is a phrase ordinarily in the data language, but *used* thus only when it describes facts about the maze. In this sentence, however, it occurs in a clause following the word *as*, where the whole sentence of which it is a part says something, not about the maze, but about the *rat's state* (of recognition). *Already entered* could be considered data language only by virtue of its antecedent, *A*, a place-term. Suppose, to take an example Tolman is fond of himself, the rats were "deceived" by the use of mirrors, or by making the "true" path resemble the other end of a detour. The rats should then avoid it, according to the hypothesis. In characterizing, for such a "deception case," either the animal's expectancy, or his actual response (i.e., avoidance) there would be no difference from similar characterizations in the Buel case; in terms of either the *behavior* or hypothesized "inner events," the psychological situation would be identical with that in which the rat is confronting an actual detour. Hence, as a psychologist, Tolman must say the same thing about the rat's state in both situations. But this shows that "... already entered" cannot here be data language referring either to (1) the maze, or (2) the animal's previous reaction, for this phrase indicates (in the Buel experiment) a fact, and in the deception case, a falsehood.

Consider a sentence of the form, "As the rat pauses at choice point 4, he expects [clause or phrase consisting of words denoting maze parts, incentives, pathways, etc.—all individually being data words]." This is intended to characterize the rat's state at the moment of making choice 4. Now, in case *E* has in fact removed or destroyed the remainder of the maze (maintaining adequate experimental control of sensory cues) the rat's state must be presumed to be as yet wholly unaffected by this fact. Hence the complex sentence (which is "about" the state of the rat) is not dependent for its correctness upon the correctness of its subordinate clauses which themselves assert (or, as in the above case of a past participle, directly imply) something about

the maze or the rat's previous history. As the logicians would say, Tolman's complex cognition-statements are not themselves truth-functions of their components.

In this connection, it is interesting to note a "philosophical" point about Tolman's system which some of its opponents have perhaps dimly perceived. Tolman himself explicitly disavows any dualistic reference for his "freshly defined" words, and has for over thirty years insisted upon his consistent behaviorism. We do not mean even to suggest that he is anything else, either consciously or unconsciously. Nevertheless, there is a peculiar sense in which his formulation has, willy-nilly, a certain affinity with the dualistic or, as he prefers to call it, "mentalistic" scheme. This arises from the feature we have just treated. For many thinkers of both past and present (e.g., Brentano), it is *intentionality* that ultimately defines the realm of "mind." Those cases of psychological description which require the use of subordinate clauses, following words like "believe," "know," "expect," such that the complex sentence is not a truth-function of its component propositions, are still a source of difficulty for philosophers basically sympathetic to the behaviorist program (cf. 11, 83). When we commit ourselves to speaking of the rat's "expectancies" rather than his "habits," we are likely to find ourselves involved in the problem of *reference, intention, or aboutness* (as in the above example) whether we like it or not. We do not suggest, of course, that such involvements are a necessary consequence of *all* formulations of cognitive theory; the sketch of formalization below seems to us free of it. But the danger is greater than in a system which is couched wholly in response language. "Learning *to* ..." is intrinsically less referential in its stress than is "learning *that* ..." The reluctance of some more suspicious psychologists to take Tolman's behaviorist protestations at face value may be subtly related to this linguistic fact.

In discussing Tolman, one needs a special terminology for this situation. We have a phrase whose components (words) ordinarily occur in the data language and there refer to parts, aspects, or relations

of the environment; if this phrase sometimes occurs in a grammatical context following behavior-words such as “know,” “expect,” “infer” (with the connective *that*), we shall say the phrase is in the “quasi-data language” when so used. We conclude from the analysis of the Buel example that the clause “that it can be recognized as but the other end of *A*, already entered” must, for a consistent behaviorism, be treated *as a whole*; that is, the sub-divisions such as “end of *A*, already entered” cannot be considered satisfactorily defined *in this context* just because they contain data words and thus *appear* to designate objects (maze parts) and events (movements). Such a phrase must be behaviorally defined. If such definitions cannot be found in Tolman, the entire phrase remains unclear in spite of the fact that all of its *elements* also occur in the data language or are readily reducible to it.

An example of the second approach to testing reducibility is found in the definition of *Inference* as given in Tolman’s glossary, viz:

*Inference.* One of the three moods of sign-gestalt-expectation (see expectation). The other two moods are perception (*q.v.*) and mnemonization (*q.v.*). In inference commerce with the sign-object only has ever occurred before. Nevertheless (perhaps because of past experience with “relatively similar” situations, or because of pure creativity), the organism is led to invent the sort of signified object and sort of direction-distance relations to this signified object which will result from commerce with the given sign-object.

The action of such inferential sign-gestalt-expectations is probably the fundamental feature in inventive learning (*q.v.*).

Inventive ideation (*q.v.*) is to be conceived as a special, sophisticated, and recondite form of inference as just defined (p. 446).

The negative part of this definition is clear (“commerce with the sign-object only has ever occurred before”) but we doubt it is an exact rendition of Tolman’s intent. The rat may have had contact with *both* the sign and significate. For example (see Fig. 3), a rat is allowed to run repeatedly from *A* to *B*<sub>1</sub> and also from *A* to *B*<sub>2</sub>; subsequently he is run from *B*<sub>2</sub> to *B*′ which contains a demanded object. If his demand for

that object is now raised and he now goes to  $B_2$  from  $A$ , Tolman would presumably consider it an “inference,” although the rat has had commerce with all significates, only not in the inference-revealing *sequence*.

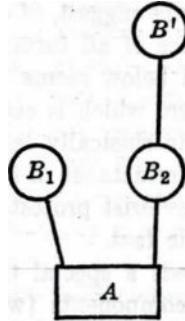


FIG. 3

As for the positive identification of *inference* in the Glossary, note the clause “... the organism is led to *invent* the sort of signified object....” The word *invent* does not appear as a defined term in the Glossary, but “inventive ideation” does. (Note: No verb forms appear in the Glossary; the rules for the use of verb forms, given the noun, are not indicated.)

*Inventive ideation.* The type of ideation (*q.v.*) called “inventive” is that in which the ideational “runnings-back-and-forth” involve, in addition to the alternative and succedent routes of a given means-end-field which the organism has already actually overtly been over, new routes or features, never as such actually experienced by the organism. In inventive ideation these new routes are ideationally extrapolated. Such extrapolation is to be conceived as brought about by behavior adjustments (*q.v.*) to those portions of the field which are already grasped perceptually or mnemonically plus the presence in the organism of a certain amount of creative instability.

The key phrase “ideationally extrapolated” suffers because “extrapolated” is not given in the Glossary. Ideation is:

*Ideation.* The more recondite form of conscious awareness (*q.v.*), in

which the sampling of alternative or succedent means-end-possibilities occurs by virtue of mere behavior-adjustments (*q.v.*) to runnings-back-and-forth.

“Recondite” is not in the Glossary, although the dictionary definition is possibly sufficient; “conscious awareness” reduces to “the process of sampling by running-back-and-forth in front of environmental object.”

Thus, if we wish to reduce this statement: *The rat infers that path P leads to food*, with the aid of the Glossary, we arrive at the following statement:

“The rat extrapolates by a more recondite form of running-back-and-forth in which the running-back-and-forth is done by a non-overtly-observable surrogate for running-back-and-forth and relative to environmental objects not then and there sensorially present, these objects being food and the path *P* leading to it.”

If we use the device of substituting a symbol for words or phrases that are neither in the data language nor in (theoretically neutral) words of ordinary English, we get:

“The rat ( $\Psi$ ) by a more ( $\Phi$ ) form of running-back-and-forth in which the running-back-and-forth is done by a non-overtly-observable ( $\Pi$ ) for running-back-and-forth ( $\Delta$ ) environmental objects not then and there sensorially present, these objects being food and the path *P* leading to it.” It is unclear whether a modifier such as “non-overtly-observable” retains any utility however precisely defined, when it refers to an unknown. We have substituted symbols for several words which do have, ordinarily, clear enough meaning, but whose meaning in this context is certainly not clear. Thus, “relative to” is reasonably unambiguous if actual running-back-and-forth is its referent, but since the running-back-and-forth referred to as “ideational” is metaphorical only, the phrase “relative to” loses the usual spatial or directional referents which give it definite meaning.

It is instructive to refer to the Glossary for the reduction of sentences which contain the word “expect.” Thus, to know what the

sentence: *The rat expects food at the end of path P*, means, we start by looking for “expect” in the Glossary. “Expect” is not given in the Glossary; however, we do find:

*Expectation.* An expectation is an immanent cognitive determinant aroused by actually presented stimuli. An expectation probably always actually occurs as the expectation of a total sign-gestalt (*q.v.*). But for purposes of analytical discourse there may be abstracted out for separate consideration, within such a total sign-gestalt-expectation, discriminanda expectations, manipulanda expectations—as to the sign-object (*q.v.*), i.e., means object (*q.v.*); as to the signified-object (*q.v.*); i.e., goal-object (*q.v.*); and as to the means-end-relations (*q.v.*), i.e., direction-distance correlations (*q.v.*) between the former and the latter.

There are three fundamental moods of expectation, viz., perception (*q.v.*), mnemonization (*q.v.*), and inference (*q.v.*).

There are no indications of behaviors in this definition. However, the definition does refer us “upward” in the definitional hierarchy to:

*Cognition* (cognitive). A generic term for one of the two classes of immanent determinants (*q.v.*) of behavior. A cognition (a means-end-readiness (*q.v.*) or an expectation (*q.v.*)) is present in a behavior insofar as the continued going-off of that behavior is contingent upon environmental entities (i.e., types of discriminanda, manipulanda, or means-end-relations) proving to be “so and so.” And such a contingency will be testified to whenever, if these environmental entities do not prove to be so and so, the given behavior will exhibit disruption (*q.v.*) and be followed by learning.

Thus, an expectation occurs as the expectation of a total sign-gestalt, that is, discriminanda- and manipulanda-expectations *as to* the sign-object, i.e., means object and signified object, and the means-end-relations (direction-distance correlations) between these two. But in the definition of an *expectation* taken literally, all references are to the stimuli involved, and these references involve a serious ambiguity because, although we generally consider references to stimuli as being in (or easily reducible to) the data language, the “stimuli” referred to in this definition of an expectancy are not the “stimuli” of the data

language at all, but rather the object of the expectation itself (intentionality again!). Insofar as this basic ambiguity exists, and conduces to an unjustified feeling of security in the acceptance and use of such a definition, we feel that this usage of “stimuli” should be understood to belong outside the data language, and in what we called above the “quasi-data language.”

A related example, pointing out the easily overlooked failure of these concepts *in use* to conform to their formal, Glossary definitions, may be found in Tolman’s discussion (128, p. 71) of the Elliott study. Water-deprived rats had run a maze to water, and error reduction was considerable. Deprivation and incentive were then shifted to food, and on the day of the shift there are increases in the time and error curves, with rapid recovery on subsequent days. The bump is explained as due to the “old cognitive expectation of water” (p. 72). A superficial and sympathetic reading leaves this explanation seeming very plausible. But we have just seen that the “operational definition” of an expectation of water amounts to saying that if the rat does *not* find water, the behavior breaks down; the “content” of a cognition is that state of the environment the *failure of which* produces a disruption of behavior. Since here it is instead the demand variable which is manipulated to produce the disruption, there is no rigorous basis for the use of the disruption criterion to establish the “cognition.” The explanation in terms of the *expectation of* a non-demanded goal object has a certain common-sense appeal; but it requires common-usage connotations, exceeding those of Tolman’s formal definition, to do the explaining job it is given here.

Although these constructs have been given unsatisfactory formal definitions, it is not correct to infer that the constructs themselves have no definitions or lack all utility. They do lack definitional precision. In such a case, definition is intimately tied to the theoretical use made of the concepts in derivation. A further discussion of the definition problem appears below, in the context of our treatment of derivation.

### *Principal Constructs*

In any listing of the principal intervening variables which Tolman has set forth, one encounters the difficulty that he has renamed some of these variables from time to time as he felt it appropriate in the context of discussion, more felicity of expression, etc. It seems safe to say, however, that if we ignore the *names* and observe the part played in the theory by the several variables, there has been little change since the publication of *Purposive Behavior in Animals and Men*. The following list is a collation of intervening variables mentioned there, and in the articles "Determiners of behavior at a choice point" (1938) (141) and "There is more than one kind of learning" (1949) (151). In each case, the anchoring variables on the independent variable side are noted wherever Tolman has himself stated them.

**Demand.** Functionally related to the independent variable *M* (Maintenance Schedule).

**Cathexis (1949) or appetite (1938).** A joint function of demand and certain properties of the object cathected, the goal-object.

**Expectancy (1932; 1949) or hypothesis 1938).** Functionally related to  $\Sigma OBO$ , a frequency notion, but most recently related also to motivation (1949) though *not* necessarily, he says, to need-reduction. There are three moods (or, variously, *modes*) of expectancy: perception, inference and mnemonization.

**Means-end-readiness (1932).** This variable is not clearly anchored; however, there seem to be two classes: a means-end-readiness may be an innately given tendency to make certain responses in given environments under various demands; in addition, it may refer to ("field-cognition modes"; 1949) a second-order disposition to acquire such first-order dispositions (3, pp. 472-481) as perceptions, mnemonizations and inferences.

**Differentiation (1938).** Functionally related to stimuli; the reference is to experimental facts more generally called *discrimination*. The concept seems not to warrant separate status; it is an instance in which

a parameter of a function describing another intervening variable, the drawing apart of expectancies, has been given separate status as though it were another “variable.”

**Motor skill (1938) or motor patterns (1949).** A dispositional concept relating to the organism’s motor potentialities.

**Biases (1938).** Functionally related to *P*, the pattern of preceding and succeeding maze units. As stated previously in the discussion of the independent variable *P*, the utility of giving these separate status from expectancies (as a function of  $\Sigma OBO$ ) is not clear.

**Equivalence beliefs (1949).** These differ from cathexes in that the object of a cathexis is actually ingested, copulated with, etc., while the object of an equivalence belief is a sort of “foyer,” anteroom, or place where cathected objects are found. They would seem to play a role somewhat comparable to that of “secondary reinforcers” in *S-R*-reinforcement theories.

**Drive discrimination (1949).** These would seem, on the independent variable side, to be anchored to the same variables as are cathexes; on the dependent variable side they seem to be undifferentiable from them. Thus, to say that a rat “knows that he is thirsty rather than hungry” presumably means among other things, that he will go wherever water is, engage in consummatory behavior with respect to water, etc. The category of equivalence beliefs means that water jars and nozzles will be cathected by him; together these would seem to cover all of the *data* that are covered by “drive discrimination.” The drive-discrimination variable specifies a qualification on the cathexis statement by asserting that all demands do not indiscriminately affect all cathexes.

### ***Relations Assumed Among Constructs***

These constructs, as enumerated above, have not been formally related to one another; if they had been, the occasional difficulties in discriminating the differences between them might vanish either

through the additional definition which the relating itself would provide, or through a redefinition by Tolman himself if the relating reveals that some of the variables are functionally indistinguishable.

There is available (128, p. 407) what appears to be a mixed causal-functional chain of variables leading to a final response occurrence. The diagram is incomplete for the present purpose, since it does not include all the proposed intervening variables. This chain runs as follows:

(1) Starting with Stimulus Variables, these *lead to* discriminanda and manipulanda-expectations, related to the organism's capacities (capacity variables: discriminanda and manipulanda capacities) which function here as parameter values;

(2) The discriminanda-manipulanda-expectations thus established lead to a Sign-Gestalt-Expectation, depending parametrically on the organism's means-ends-capacities (that is, capacities for perception, mnemonization and inference). It is not clear here whether the Sign-Gestalt-Expectation is supposed also to depend on the means-end-readinesses, which include not only the "endowment" (capacity) but also the past-training values of perception, mnemonization and inference, but presumably it does.

(3) This Sign-Gestalt-Expectation leads to "Running back-and-forth or adjustment to Running-back-and-forth" (an earlier form of *VTE?*) depending parametrically on the *O*'s consciousness-and-ideation-ability.

(4) As a result of this Running-back-and-forth (*VTE*) the *O* provides himself with  $S_2$  (new stimuli, presumably both exteroceptive and interoceptive) which lead to new manipulanda-discriminanda expectations, thence to new sign-gestalt expectations, with the same parameters affecting these linkages.

(5) Throughout the chain, the Demand variable has been present; "These depending demands control the whole line of the  $S \rightarrow R$  process" (p. 407). An initial physiological state leads to a "Demanded Type of Goal" (presumably this means *demand for a given* type of goal) depending on means-end-readinesses, innate and acquired, and also demand for subordinate (equivalence belief) goals.



(6) The all-important linkage to response is handled rather casually as follows (p. 407): “[These demands lead the organism] to respond to the given *S* as presenting an appropriate means-object.”

It has been repeatedly pointed out by Tolman’s critics that this sort of diagram is really not a *system* or a proposed set of laws, but rather a framework *for* laws, in which we are told *what is related to what* in the construct system but not much (if anything) about the character of the relations.

### ***Methodological Characteristics***

**Explicitness of formulation.** Until the publication, in 1949, of the article, “There is more than one kind of learning” (151), the explicitness of Tolman’s system was limited to a statement of the *kinds* of variables and kinds of axioms that he intends to use (see *Purposive Behavior in Animals and Men*, pp. 372-374, for his list of “modified and new laws” for conditioned response, trial and error, and inventive learning). The 1949 article discusses the laws of acquisition, de-acquisition and forgetting of cathexes, equivalence beliefs and (field) expectancies. The explicitness of these laws may be estimated by a consideration of them separately.

For *cathexes*, acquisition depends upon the adequacy of the object for physiological need reduction, and it is guessed that “numbers of repetitions and amounts of need reduction per repetition would, no doubt, turn out to be the two major causal variables and that the curves would undoubtedly be exponential in form” (151, p. 147). For de-acquisition, if need reduction following, say, the ingestion of food could be prevented, the cathexis for such food would weaken. Cathexes are not, however, forgotten with the mere passage of time.

For *equivalence beliefs*, primacy, frequency and intensity of need-reductions, as well as early traumatic experiences, are tentatively suggested as the principal causal factors in acquisition. There is no guess as to the condition of de-acquisition (although this is recognized as a major clinical problem; the neurotic may be thought of as having

erroneous equivalence beliefs). As for forgetting, Tolman admits that the evidence is small, but postulates that equivalence beliefs are not forgotten but have to be unlearned. It can be seen here that, within the limits of the tentativity of these “axioms,” an additional differentiation between cathexes and equivalence beliefs is suggested in the clear difference in their laws; cathexes weaken if need-reduction is withheld, but the hall-mark of the equivalence belief is the substitute acceptability of a *means* for an *end*, and hence the continued strength of an equivalence belief without need-reduction. This seems to put Tolman with Allport in the matter of functional autonomy.

For *field expectancies* (the older sign-gestalt-expectations and the newer cognitive maps), the factors of acquisition are frequency and motivation. “But this does not mean that I hold that such learning consists in the stamping in of *S-R* habits by reinforcement ... [a goal] ... probably does give a special vividness to that locus in the total field expectancy” (151, p. 151). To these must be added the perceptual, mnemonic and inference *abilities* of the organism. De-acquisition occurs when, and only when, the environment changes so that the previous expectancy is unsuitable; hence de-acquisition is an “interference” at the cognitive level. As to forgetting of expectancies, true forgetting with the passage of time does occur; particulars may be lost, in the Bartlett-Gestalt sense of simplifying and sharpening.

For *motor patterns*, Tolman points out (151, p. 146) that these “have to be included by me, since I do not hold, as do most behaviorists, that all learning is, as such, the attachment of responses to stimuli.” To include them, Tolman expresses a willingness to agree with Guthrie that *movements* are learned rather than *acts*, and the Guthrian formulation of conditioning by contiguity is accepted with a protest that movements which are learned are “embedded in a larger goal-directed activity.” These movements, when learned, are available to the organism for trying out in new situations. Great caution would seem to be required in the interpretation of these assertions; if they are taken quite literally, a good case could be made that these “laws” of

movement acquisition involve a striking concession, if not an essential capitulation, to the peripheral (as opposed to the central) entailments of an *S-R* or *S-R*-reinforcement theory. Unfortunately, Tolman has not been very explicit as to the limits of this agreement with Guthrie. If, however, we attend to other very clear statements as to the central character of a learned expectancy, and to the spirit of the system, we note: "Cathexes, equivalence beliefs, field expectancies, field-cognition modes and drive discriminations are not, as I define them, stimulus-response connections. They are central phenomena, each of which may be expressed by a variety of responses. The actual nature of these responses is, however, also determined by the character of the motor patterns at the organism's command" (151). Perhaps an acceptable translation would be: *given* the final field expectancy, the *form* of the responses may be a motor habit. The problem of getting to the response has always been considered to be a special problem for Tolman. Part IV of this section will consider this "problem of getting to the response" and offer one kind of suggestion for its solution.

Tolman has only recently offered the above axioms; they are seen to be very tentative and imprecise. These are not, of course, sufficient grounds for criticism; the sense in which they are criticizable will be noted in the section in this chapter dealing with our own suggestions toward axiomatization.

**Quantitativeness.** Tolman's system would seem to be so constituted as to permit quantification, but the step has never been taken. Hilgard (37) has stated:

In spite of a clear outline of what a systematic theory ought to be, Tolman has nowhere attempted quantitative predictions paralleling those of Hull, so that his conjectures have not, in that sense, been put to the test. This does not mean that his experiments are unrelated to his theory. There are, in fact, many predictions, but they assert that one path will be preferred to another, that under one set of circumstances the problem will be easier than another set, and so on. The dimensional analysis that completes the function is not provided (p. 265).

It is difficult to make an assessment of the degree of quantification of this or any other theory. See Part IV, this section, for a possible “scale of quantification,” with remarks as to Tolman’s position on it.

**Consistency and independence of the postulate set.** In order really to evaluate the consistency of Tolman’s system, explicit formalization is a necessary first step. Since this has not been taken, the evaluation is impossible. Similarly, since to show that the postulates are independent would involve showing that none can be derived from the others as theorems, the proof of independence would presuppose an explicitness of axiomatization not present in Tolman. Formal treatments of the question of independence are hardly necessary at this stage of behavior theory.

**Models.** Tolman has not used models in his system-making up to 1949.

**Techniques of derivation.** The derivation of experimental results from Tolman’s system is, due to the informal axiomatization which characterizes it, exclusively informal. To illustrate this informality, it is an interesting exercise to note in a recent paper by Tolman “... the actual experiments ... out of many which I have selected to report ... which seem especially important in reinforcing the theoretical position I have been presenting” (150, p. 193), and to examine critically his method of relating these *to* the theory. He treats of five classes of experiments:

(1) *Latent learning.* In discussing the results of the Blodgett experiment, Tolman says:

It will be observed that the experimental groups as long as they were not finding food did not appear to learn much. (Their error curves did not drop.) But on the days immediately succeeding their first finding of the food their error curves did drop astoundingly. It appeared, in short, that during the non-rewarded trials these animals had been learning much more than they had exhibited. This learning ... Blodgett called “latent learning.” Interpreting these results anthropomorphically, we would say that as long as the animals were not getting any food at the end of the maze they continued to take their time in going through it—they continued to enter

many blinds. Once, however, they knew they were to get food, they demonstrated that during these preceding non-rewarded trials they had learned where many of the blinds were. They had been building up a "map," and could utilize the latter as soon as they were motivated to do so (pp. 194-195).

Aside from the fact that it is not quite correct to say that the error curve of the latent group did not drop in the latent phase, one could note that the actual "derivation" of latent learning phenomena *from* Tolman's theory here is simply not given. The tie-up is made only by pointing to it as a presumed instance. His earlier negative use of Blodgett's results, as a criticism of the law of effect, was clearer (128, p. 343).

(2) *Vicarious trial and error, or VTE:*

But what, now, is the final significance of all this *VTE*ing? ... My answer is that these facts lend further support to the doctrine of a building up of maps. *VTE*ing, as I see it, is evidence that in the critical stages—whether in the first picking up of the instructions or in the later making sure of which stimulus is which—the animal's activity is not just one of responding passively to discrete stimuli, but rather one of the active selecting and comparing of stimuli (p. 200).

Here derivation is pursued by the device of opposing such vague, undefined terms as "passively" and "discrete stimuli" to such phrases as "active selecting and comparing of stimuli." It is hardly possible to view this as acceptable deduction of experimental consequences, even within an avowedly incomplete theoretical structure.

To get from the axiom: *Organisms form maps*, to the theorem: *VTE occurs*, it is not sufficient, or even clearly relevant, to refer thus loosely to the contrasted *S-R* formulation. However, since it *is* referred to, one may question the incompatibility between the statements:

- (a) Jumping to a white card is strengthened by a reinforcing consequence; jumping to a black card is weakened by its absence; and
- (b) Animals look at the stimuli several times before they jump.

As a matter of fact, Spence (95) has shown how "non-passivity," or

“paying systematic attention to the relevant stimuli,” may be derived as a consequence of reinforcement theory. We are not concerned here, of course, to defend Spence’s derivation. Even if it were rejected, the point is that Tolman seems to take as somehow obvious the relation between his (or Spence’s) theory and the facts, letting such words as “passively” and “significant” serve as substitutes for an actual derivation. It may be that statements (a) and (b) above *are* incompatible in some way not clear to us; but surely their incompatibility needs to be exhibited by some line of argument.

(3) *Searching for the stimulus.* The experiment referred to is an unpublished one by Hudson in which rats failed to exhibit an avoidance response to a stimulus which disappeared coincidentally with the administration of an electric shock. In commenting on this, Tolman says:

I feel that this experiment reinforces the notion of the largely active selective character in the rat’s building up of his cognitive map. He often has to look actively for the significant stimuli in order to form his map and does not merely passively receive and react to all the stimuli which are physically present (p. 201).

The alleged inference seems to hang chiefly on the vague and emotionally connotative word *passively*. The impossibility of deriving this “searching” on *S-R* principles also needs to be shown; and there is no move made toward showing it.

(4) *The “hypothesis” experiments.* Reference is made to the non-random response sequences exhibited by Krechevsky’s rats when he presented them with an unsolvable problem in maze running. Again, Tolman does not relate this phenomenon to the theory.

(5) *The spatial orientation experiments.* None of these is explicitly derived in the article. However, the point of the spatial orientation series in supplying an evidential basis for a cognitive map view (and against *S-R* views) is somewhat obscure because of the failure of theorists on both sides to define clearly what they intend by “a response.” This will be discussed at length below.

Two things may be said in respect to these five classes of experi-

ments as supports for the cognitive map point of view: First, those which are offered as embarrassments of law-of-effect or contiguity formulations cannot be considered as arguments *for* the cognitive map formulation unless the competing systems are shown to be mutually exclusive *and* exhaustive. Secondly, there is no evidence of formal derivation in Tolman's discussion of them. It does seem as if these results might flow as consequences from "some such theory," and the failure to be able to show whether they would do so is due to the method of definition of concepts and the informality of the axiomatization.

It is appropriate now to refer back to the question of the definitions of the concepts, whose clarity has not increased much by noting their use in derivation. The argument might be made that, for example, the terms for which the Greek letters were substituted above are *implicitly* defined. Thus, when a critic asks for a definition of a certain term or expression which occurs in a scientific theory he may be answered simply by exhibiting all of the sentences of the theory. The "meaning" of a construct is disclosed by seeing its role in the system. Philosophical workers in the field of logical analysis are increasingly aware of serious and involved problems in this area (9, 34, 82). To the extent that words occurring in the theorems have empirical coordinating definitions, and if the theorems really do follow from the axioms, then all of the words in the axioms are, so to speak, "given empirical meaning." Therefore, in an admittedly incomplete system the *truth* of the axioms is not only doubtful, but even the *meaning* of the terms partakes of a certain vagueness. Such a set of terms is said by Hempel to be "partially interpreted." There are many difficulties and unsolved problems here, and psychologists cannot afford to be dogmatic when the methodologists are in such doubt.

Recognizing this unsettled condition of the general issue, it seems admitted that the meanings of such implicitly defined terms, are, so to speak, nailed down by piling up the theorems derived from them. When a word, e.g., *hunger*, is found in several contexts, it is possible

to “take a fix” on it, to carry out what Feigl has called “triangulation in logical space.” The defining of terms by exhibiting their formal roles presupposes, however, a formal role to exhibit; hence, discussion of implicit definitions in Tolman’s system has necessarily been delayed. The Greek letters in our expansion of the “Inference” construct might be justified as introduced by implicit definitions via the system as a whole, were it explicitly axiomatized. But, as has been shown, these derivations are “informal,” when they are not simply absent! In such a system it is hardly possible to speak of implicit definitions since the formal role of a construct cannot very well be exhibited informally. We conclude that Tolman’s formulations do not permit any satisfactory implicit definitions for the “operational” ones he officially advocates.

### ***Empirical Content and Adequacy***

**Range of data purported covered.** Tolman’s is obviously intended to be a “complete” psychology of “docile” behavior. In all fairness, it must be said that he is under no illusion that he has, indeed, completed it. It has been noted previously that rats in mazes have nearly exclusively been his source of empirical support, so that at least he could be considered as intending to develop a fairly complete theory of maze behavior. Whether the laws of learning as “laid bare” in, or by, the rat, will ever constitute a complete psychology in a broader sense is a question that cannot be answered until its own laws are reasonably complete; but this is applicable to any theory and is hardly uniquely remarkable of Tolman.

**Specificity of prediction demonstrated.** Insofar as Tolman’s is not a highly quantified system, predictions from it have been, perhaps exclusively, of a *presence or absence, right or left, more or less, faster or slower* sort. Insofar as the axioms and derivations are informally stated, it is not always clear whether the predictions that have been made follow from the theory or whether they are, as some are admittedly, guesses made anthropomorphically or on the basis of common sense.

**Obvious failure to handle facts.** Since the derivation of experimental results has been characterized by extreme informality, gaps in the theory are difficult to detect. There is, at least, a vocabulary with which to discuss most of the data of learning, although the quantitative vocabulary is lacking as the theory now stands. The inadequacies of definition of the principal constructs (as pointed out earlier) suggest that more complete operational definition of the concepts would make them unsuitable for some of the explanatory sentences in which they now stand. A more flexible method, and one more likely to preserve the connotations of the constructs as Tolman now uses them, would be to define these terms implicitly, in which case a test of the adequacy of the theory with respect to the data would depend upon the explicitness of the axiomatization. But as neither of these steps has been taken, the range of applicability of the theory is nearly impossible to assess. The temptation to over-estimate the utility of the concepts is heightened by the non-technical connotations that accrue to such words as “expect” and “infer.”

There are certain guarantees built into the system which may be used to account for failures of prediction. Thus, *emphasis*, as a relatively undefined property associated with motivation, may be used to account for the failure of the rats in one type of latent learning study (called Type 4 below) to learn where a non-valenced goal-object is to be found; or “fixation” of responses by overlearning, or emotion, may be used to account for failures of place-learning. We do not mean to suggest that these concepts are intrinsically without merit; but until they have been more carefully defined and quantified, there is no clear restraint on their *ad hoc* usage.

### ***Programmaticity***

The most serious obstacle to an evaluation of the cognitive learning theory of Tolman is the unspecificity of its axiom set. Tolman has shown a tendency to restate the system by revising its vocabulary (i.e., “sign-gestalt-expectations” become “cognitive maps”) and by recount-

ing illustrative experiments (often mainly because they are disquieting to the law of effect theorist, though only thus deviously related to the affirmative side of the cognition theorist's argument). The programmatic nature of this system is most clearly suggested by noting that in 1932 *Purposive Behavior in Animals and Men* lists (pp. 372-374) only the *kinds* of laws that will be needed in a purposive behaviorism; and in 1949 he says, "And although, as usual, I have been merely grammatic and have not attempted to set up, at this date, any precise systems of postulates and deduced theorems, I *have* made some specific suggestions ..." (151, p. 154).

### III. ABILITY OF TOLMAN'S THEORY TO ACCOUNT FOR FACTS NOT PREDICTABLE FROM COMPETING THEORIES

If we ignore the question of whether such facts are accounted for by Tolman's theory as it stands, we can note that the data of the *latent learning* studies are the kind which Tolman intends to account for, and for which he believes *S-R* and *S-R-reinforcement* theories are inadequate. The experimental designs which are used to test for latent learning are very diversified; Thistlethwaite has reviewed this literature (106). In general, latent learning is revealed in an abrupt, or somehow discontinuous, change in response tendency when drive and incentive are first irrelevantly matched (the latent learning period) and then relevantly matched (the test period or trial) in a given experimental setting. "Relevant" and "irrelevant" are definable in terms of maintenance schedule operations, behavior of the organism in the presence of the incentive, etc. "Abruptness" may be given statistical definition, as a significant shift from the random or systematic response tendencies before the drive-incentive matching experience (as in the T-maze studies, the behavior of the rat running the maze hungry for the first time after previous exposure to a food incentive in the maze while satiated for food; or as in the Blodgett design, the reduction in cul-

entries when rats first find food in the maze after having run the maze while hungry but without food in the goal-box).

Part of the question with respect to the latent learning studies is the extent to which they do reveal inadequacies in the *S-R*-reinforcement formulations. From this standpoint, certain of the designs seem to us not to be determinate, as will be seen below. In this regard, it may be noted that the reinforcement theorist's use of the  $r_G$  construct, at least its use according to a broad definition in which no peripheral event is entailed, can apparently mediate any outcome which a cognition theorist would call an "expectancy." The  $r_G$  construct *with* a peripheral event entailed is equally powerful (although severely strained for the Blodgett and free maze exploration types), but then the burden of proof of the peripheral activity presumably falls on the user of the concept.

Five categories of latent learning studies are noted:

1. *The Blodgett type*. Drive is strong and unmanipulated (by maintenance schedule) throughout; the incentive is introduced later in a series of maze runs. Latent learning is revealed by the error record on the trial subsequent to the first feeding in the maze if the drop in errors is larger than the first, or the largest single, drop in the error curve of a control group which has been fed from the first run, or by a steeper drop in errors from the same level on the ordinate. There have been nine such studies reported; of these, seven have been interpreted as positive [Blodgett (2), Elliott (23), Herb (35), Simmons (89), Tolman and Honzik (162), Wallace, Blackwell and Jenkins (184), and Williams (186)]. Two are negative [Meehl and MacCorquodale (65), Reynolds (76)]. The "latent" groups in both these studies showed considerably more error reduction during the latent phase than Blodgett's did.

The Blodgett experiment is extensively cited as *the case* for which *S-R*-reinforcement theory is inadequate, and it is often used to display the difference between *S-R*-reinforcement and cognition theories. Hence it is important to examine its design and the results it provides to determine whether it is, indeed, an *experimentum crucis* in any respect.

The core of the argument from the Blodgett design concerns the drop in errors following a reinforcement introduced after a series of less strongly rewarded maze runs. It is not the fact of, but the size of, this error reduction that is crucial. *S-R*-reinforcement theory predicts a drop in errors following the first rewarded run; but cognition theorists assert that this drop is actually greater than reinforcement principles can account for, and thus that in the pre-reinforcement runs “learning that didn’t show” was taking place. As an estimate of the magnitude of error reduction to be predicted from one reinforcement, the cognition theorist points to the error drop occurring after the first reinforcement of a *control* group, which is fed on its first run. The comparison may also be made against the largest single drop in errors shown by the control group, or against the drop made by the controls from the same point on the ordinate. The third of these comparisons has the most methodological soundness in terms of learning curves, and we shall confine our discussion to it as an example of the difficulties met in interpreting the findings of the Blodgett experiment. The important question to answer here is whether the reinforcement operation by being performed not on the first, but on, say, the seventh run would actually be presumed by the *S-R*-reinforcement theorist to have the same effect on error scores. By his theory the effects of a given reinforcement upon the habit strengths of the two groups may be equal, but the drop in the error scores ( $sE_R$ ) may still be unequal if other variables which affect the *utilization* of habit strengths are unequal at the times involved in the comparison.

Two major variables specified by *S-R*-reinforcement theory as affecting the probability of response when habit strength is constant are (1) drive and (2) competing  $sE_R$ 's. If in the Blodgett experiment reductions in errors following a reinforcement are to be used to measure the resulting shifts in *habit strength* (not  $sE_R$ ) the structure of the experiment should be inspected for its guarantee that the conditions of drive and competing  $sE_R$ 's are equal for the two groups at the times of the comparisons. “Guarantee” means showing that conditions

for the two groups should theoretically be equal, except for the random factors which enter the error term of the significance test.

The burden of this proof of equality rests with the user of this design if he intends to demonstrate the inadequacy of reinforcement theory to account for *as much* learning as is observed. The *S-R*-reinforcement theorist need not, actually, show that these other factors are *in fact* unequal in this design, if he can show from findings in other experimental settings that the different histories of the two groups *might have* produced systematic differences in either drive or competing response tendencies by the times the comparisons are made, and that these differences would be in the direction which favors a larger drop in errors by the experimental group. If this can be done, the cognition theorist's argument that these histories have produced differences unaccounted for by reinforcement principles is vitiated. Only to the extent that in the Blodgett design "equal errors" implies "equal habit strength" is the cognition theorist's argument *from* the Blodgett design valid; to the extent this equivalence does not hold, the argument is invalid. The Blodgett type of experiment is cited by cognition theorists as inexplicable by *S-R*-reinforcement theory. In such an argument, one must obviously allow the *S-R*-reinforcement theorist *all* of his theory when he is asked to account for the results. Consequently the burden of proof upon the cognition theorist includes a showing that the conditions for any comparison within the framework of *S-R*-reinforcement theory have been met by the experimental design.

It seems doubtful that he can do so. When the comparison in the Blodgett design is between the Experimentals' post-reward drop and the drop in the Control's curve from the same point on the ordinate which characterized the Experimentals on their first rewarded run, what should be the *S-R*-reinforcement prediction? *S-R* theory predicts that the effects of reinforcement would be equal for the two groups only if the two groups are shown to be equal in *three* additional respects: (1) the initial  $sH_R$ 's must be the same; (2) the increment to

$sH_R$  must be the same; (3) following the reward with the resulting and still equal  $sH_R$ 's, drive and competing  $sE_R$ 's must be the same so that the equal  $sH_R$ 's will yield equal post-reward error scores, hence equal drops in the error curve. Let us examine these three required assumptions.

Consider first the assumption of equality of the  $sH_R$ 's at the point of comparison. In the cognition argument this is claimed from the equal height of the ordinates just *before* the experimental group is reinforced. But equal performance implies equal  $sH_R$ 's only, again, if (1) the drive states of the two groups are equal, and (2) competing responses are equal. There is some reason to doubt that in the Blodgett experiment the drive states of the two groups were equal at the times the "corresponding" reinforcements were given. If the comparison made is with reference to an ordinate reached by the controls at any time subsequent to their first maze run, *S-R*-reinforcement theory implies some degree of drive inequality favoring the controls, on the basis of conditioned motivation. Exteroceptive maze cues should become elicitors for some of the components of hunger-drive as a result of the feeding experiences in the maze situation, so that after one or more food-rewarded runs the controls should be psychologically "hungrier" than the experimentals on the same maintenance schedule who have not, however, been getting fed in the maze. This concept of conditioned drive was explicitly set forth in Hull's 1949 postulate set (39, Postulate III, Corollary i, p. 175) and appears in his posthumous book (40, Postulate III, Corollary i, p. 6). Earlier (1941) but less formally, the notion was developed and extensively utilized by Miller and Dollard in their application of *S-R*-reinforcement theory to the human social case (68, Chap. IV and *passim*). Once the idea that there are stimulus and response components of drive is admitted, the occurrence of drive-conditioning flows readily as a consequence of the postulates proposed by Hull in his 1943 book (38). Several experimental approaches to this concept are possible, one of which we have taken in an experiment modified from Blodgett's (67). We showed

that a feeding experience in a maze-like situation, *not* following a maze run and *not* at the goal-box end (or in its direction), produced a significant drop in errors and faster running, with no further maze exposure to yield an increment in habit strengths (or cognitions, for that matter). This finding would seem to support a suspicion that Blodgett's experimentals must have lacked some of the evocative drive of the controls, due to the lack of opportunity for acquiring such conditioned motivation. In a series of studies utilizing a different experimental setting, Seward, Datel, and Levy (86) attempted to test our hypothesis, with equivocal results. *If* the experimentals, on their last pre-food-reward day, are less motivated than the controls on their comparison day, a matching of ordinates (reaction potentials) implies a systematic *un*matching of habit strengths. That is, at a comparable level of performance the experimentals (being less motivated) must have more habit strength favoring the correct path than the controls. The food-reward given for the first time at the *end* of this run makes available, on the following day, a heightened drive factor. According to Hullian theory, this increment acts mathematically as a *multiplier* of any habit-strength differences between true path and culs, and (under a suitably chosen range of parametric assumptions) should produce a steeper drop than that given by the habit-strength increment at some post-initial stage of the acquisition process.

During the latent phase, the experimentals are accumulating  $sH_R$  favoring the true path. Two replications of the Blodgett experiment (65, 76) show considerable error reduction before the first feeding of the experimentals; Tolman and Honzik's curves show the beginning of such a reduction. It must, of course, not be forgotten that this first rewarded run *also* contributes an increment to habit strength. On our assumption that the experimentals' "unmanifested" habit strength exceeds that of the controls when the latter's ordinate was comparable to that of the experimentals immediately pre-reward, the  $sH_R$  increment yielded by one food-reward should be less for the experimentals (due to the decelerated form of the  $sH_R$ -acquisition function). This

complication operates in opposition to the effect just discussed. Whether it would be sufficient in amount to obscure the hypothesized effect entirely involves several rather complex quantitative questions, not readily (if at all) answerable within the Blodgett design itself. Avoiding rash assumptions about these matters, it seems legitimate to argue that the drive-conditioning hypothesis must be given serious consideration as one of the contributors to the Blodgett effect; and that, in the light of our present ignorance of the parameter questions involved, it cannot be excluded as a possible *S-R*-reinforcement explanation of the phenomenon.

But even this does not exhaust the complications which the design generates within a Hull-type theory, and, presumably, for any form of *S-R*-reinforcement theory made sufficiently explicit to predict for the maze case. In the preceding discussion we have treated the habit strength and drive multiplier as if they were associated with some abstract, "generalized" choice-point; so that the quantitative arguments relating changes in these variables to the observed datum (*total errors per maze run*) would hold strictly only if the relative effects of conditioned motivation and  $sH_R$  increment were not different over the several maze units. And this is surely not the case. Having been given food-rewards in the goal-box, the controls (when manifesting any given level of "performance" defined as *total errors*) can be assumed on the basis of other maze studies to show a start-to-goal error gradient, although Blodgett does not present an error breakdown by units. This gradient should be less distinct for the experimentals, who (when at a comparable total-error performance level) have been operating on the basis of non-alimentary incentives and for whom the goal-box as such must be presumed to be less rewarding. In other words, the "total correct choices" for the experimentals on their last pre-food-reward day is very probably distributed *more equably over the six units* than is the same quantity for the controls. This situation creates a further asymmetry in the relation of the two sorts of strengthening operations being compared (drive-conditioning versus  $sH_R$ -increment

due to the reward); and the asymmetry favors a Blodgett effect. For when the drive-conditioning factor is already present (control group), the increment in  $sH_R$  given by one additional food-reward *cannot* be assumed to be equal over all six units, and in fact, in the case of units in which an error was made during the run in question, the impact of the terminating goal-box reward is a very delicate question. It might conceivably even be negative at such early stages of learning as the second or third day, before secondary reinforcement has moved back far enough to create a series of supporting sub-goals. One would, therefore, expect the  $sH_R$  increments yielded by a single food-reward for the controls to be unequally distributed over the six units, varying considerably from rat to rat depending upon just what he did (and the all-important temporal factors of his doing it!) on the run in question; but, statistically speaking, being progressively feebler as we consider units more remote from the goal.

For the experimentals, with a more equalized  $sH_R$  difference from unit to unit, the drive-conditioning operation should theoretically exert a more massive effect, since it operates mathematically as *a multiplier applied to all six units at once*. A little computation with plausible quantitative examples will convince the reader that this asymmetry between the reward as an  $sH_R$ -increaser and the reward as a drive-conditioner can, within a considerable range of parametric assumptions, make a marked difference in the performance error-drop. A fairly direct test of the above reasoning would probably require a longer maze and large numbers of animals, since the Blodgett maze is very highly patterned as to error-frequencies and the sixth unit is so quickly eliminated (even by non-fed animals) as to yield an essentially five-unit maze after very little exposure.

Consider next the third necessary assumption, that of the equality of competing response tendencies. The chief competing response tendencies involved in maze running may be subsumed under the general rubric "exploratory disposition." Whether it is appropriate to speak of the exploratory disposition as a *drive* or not we need not discuss here.

It does seem to have at least some drive properties, such as dependence upon a maintenance schedule and an energizing effect upon behavior capable of overcoming opposing incentives such as shock. Whether we speak of it as a drive or not, the fundamental operation known to weaken exploratory behavior is repeated exposure (18, 19, 20, 58, 69, 70, 71, 187) and this has been permitted to the animals in Blodgett's experimental group to the extent that their first food-rewarded run is delayed until later in the series of maze exposures. No one who has observed rats during their early exposures to a maze could dismiss the exploratory disposition as of negligible strength. Not to mention the specific experimental quantification of this "need" in the Columbia obstruction-box studies, one thinks of such informal (but common) observations as the long latency in eating shown by 24-hour hungry rats in a novel situation, and the frequent interruptions of this overlearned consummatory behavior by the competing "investigatory" responses. We have unpublished data showing that if hungry rats have been run and rewarded in the Blodgett maze with culs blocked off, the number of trials being more than sufficient to yield errorless runs as this maze is ordinarily used, unblocking the blind alleys leads nearly every rat into every cul. We have also shown (58) that the tendency to enter a cul is an inverse function of the "total past penetration" into it, quite apart from any reward associated with getting to the maze end.

These considerations show that what is customarily designated as a "learning curve" in multiple-unit maze studies is *not* a pure learning curve (either in the sense of habits or cognitions), but is a composite curve—the resultant of interaction between the underlying course of learning and the reduction ("satiation" or "negative adaptation") in the exploratory tendency. The usual maze experiment presumably gives us a very obscured indication of the building up of habits or cognitions; we do not know what the form and parameters of the acquisition function would be like for an "ideal rat," constructed so as to be a pure  $sH_R$ -accumulator (or knowledge-acquirer). We suggest, however, that

the kind of performance curve exhibited by Blodgett's experimentals, once their food-reward is introduced, is a better approximation to the idealized case of a learning process uncontaminated by emotional and investigatory interferences—a better mirror of the underlying acquisition function—than is the usual “maze-learning” curve as represented by the controls. Regardless of the stand one takes on this question, it seems fair to say that matching performance ordinates is *not*, *prima facie*, matching the underlying  $sH_R$ 's, but rather is systematically unmatching them. Here, however, the presumed (correct-path)  $sH_R$  favors the controls, since their ordinate is a resultant of the influence of the true-path  $sE_R$ 's and an opposed “exploratory”  $sE_R$  which is stronger than that of the experimentals (on a later day). The impact upon  $sH_R$  of a food-reward is thus greater for the experimentals, making the usual Hullian assumption of deceleration for the  $sH_R$ -acquisition function. Hence the decline in errors following this reward should be steeper than that for the controls starting from an equated performance level.

But there are still additional complications. Whatever the impact of the single food-reward upon the habit strengths, what of the competing exploratory  $sE_R$ 's on the *next* trial? We have suggested that the effect of the first food-reward upon the next day's choices for the experimentals may be looked upon as the “pure, unadulterated” consequence of reinforcement, now unimpeded by the interference of exploratory responses. Whether this factor works “for” or “against” the Blodgett effect depends upon further quantitative issues. If one thinks of the effective  $sE_R$  for true path as a linear resultant of the exploratory  $sE_R$ 's for cul (*and* true path!) together with the new  $sE_R$  contributed by food-reward, the *form* of the exploratory-satiation function becomes all-important. If it is markedly decelerated, this summative process works counter to a Blodgett effect; since the decrement in cul- $sE_R$  between two early trials (control group) should exceed that between two late trials (experimentals). If it is approximately linear over the interval concerned, this complication can be ignored, and the above develop-

ment predicts a Blodgett effect. If, finally, it is ogival, and the flex point lies somewhere between the abscissas determined by the equating of performance ordinates, whether a Blodgett effect were favored or not would depend upon the slopes at the two points of the ogive.

The preceding discussion treats the exploratory disposition as if its only contribution to the situation lay in its role as an  $sE_R$  for cul entry, competing with the "correct"  $sE_R$ 's. Again, even superficial familiarity with the actual character of a "naive" rat's early maze behavior shows that this is a gross oversimplification. In terms of  $S$ - $R$ -reinforcement reconstructions of maze-learning, the greater strength *and variability* of such concurrent investigatory responses as halting, reversal, standing up, sniffing in comers, freezing, sudden spurts and the like, so characteristic of the less adapted rat, should theoretically reduce the efficiency of the various mediating mechanisms (e.g., chaining effects, secondary reinforcement, "pure stimulus acts," drive conditioning) and hence attenuate the impact of a reward given at the termination of a run.

It must be emphasized that the entire preceding discussion of the probable role of competing  $sE_R$ 's contains lacunae at several critical stages due to our experimental ignorance of certain quantitative matters; and that some of the possibilities work in directions antithetical to others, although to an unknown extent. None of the behavioral tendencies invoked are purely *ad hoc*, however, being here carried over from empirical findings in other setups. Some of them have received reasonably direct support within the Blodgett situation itself. At the least, it seems fair to say that the required assumption of equality of competing  $sE_R$ 's is not *prima facie* justified, but, if anything, rather unlikely. The core difficulty seems to be that any experimental procedure which allows a rat an opportunity to establish *either* cognitions or habits in a maze will also, unavoidably, permit him to adapt emotionally and "satiare" his exploratory drive. To the extent that he has,  $S$ - $R$ -reinforcement theory predicts that the effect of one reinforcement will be, if not an exactly specifiable amount, at least greater than

that for an unadapted, unsatiated animal.

A final complicating factor in the Blodgett design is that we are unable to state definitely that the two groups are equal in unconditioned drive strength. The crucial reinforcements are given at times when the two groups have been on the feeding cycle for different periods. In Blodgett's experiment a pre-experimental feeding cycle was in effect only three days; thus the control group was first reinforced on the fourth day, while one experimental group was first reinforced as late as the tenth day after inception of the cycle. We attempted (65) to reduce the relative discrepancy necessarily involved here by extending the pre-experimental feeding cycle to *seven* days. It is surely possible that by the time of its first feeding, the experimental group, with its longer history of deprivation periods, has an accumulated deficit which makes it hungrier. It would be difficult to do, but the design would seem to require that a prolonged pre-experimental feeding cycle be used, to ensure that both groups are at the asymptote of (unconditioned) drive strength for the deprivation period used (22 hours) and the amount of food given. It seems unlikely that such stability could be achieved unless the amount of food given were carefully calibrated so as to ensure day-to-day equality of drive strength for "22-hour hungry rats." The literature on the effect on  $sH_R$  of deprivation at the time of reinforcement is ambiguous (6, 29, 44, 60, 77, 103); the functions seem not to be linear or even monotonic, but the evidence does not preclude drive level's acting as a parameter in the acquisition function. If there is a range over which this relationship is positive, and if the Blodgett design is operating within it, the influence for the experimental group opposes the action of conditioned drive, presumed (as we have seen) to be *less* for the experimentals before and at the time of its first reinforcement. But on the other hand, it suggests that the  $sH_R$  for the experimentals must *have been* lower than that of the controls on the matched run, and, therefore, that the  $sH_R$ -increment yielded by the terminal reinforcement should be larger for the experimentals due to the deceleration of the  $sH_R$ -acquisition

function. Hence this possible inequality in unconditioned drive works at two points in the intervening variable chain in favor of a Blodgett effect, and at one point in opposition.

Whatever its effect at the time of reinforcement, drive strength is clearly at issue as a relevant variable on the first post-reinforcement run, when the  $sH_R$ 's created earlier are "activated." In predicting the size of the drops for the two groups, the possibility of this difference and its direction must be taken into account—conjointly, of course, with the differences between the  $sH_R$ 's and competing response tendencies of the two groups just previously discussed.

We seem to have a plethora of contradictory hypotheses to account, reinforcement-wise, for various Blodgett-design phenomena. The fact that there are so many and some subsets of them are contradictory shouldn't reflect on them, nor does it unsettle us: the point is that the Blodgett design *presupposes* the availability of a valid estimate of the experimentals' error drop. Our welter of hypotheses shows that making this estimate is just not feasible.

Since this is a chapter on Tolman, not Hull, we have not commented upon the derivation of the Blodgett effect which appeared with Hull's last book (40, pp. 140-148). Hull's analysis has not, at this writing, been the subject of discussion by Tolman's school. It invokes the new factor of incentive motivation, or  $K$ , in a manner mathematically analogous to our use of drive-conditioning. For a commentary on the  $K$ -component in Hull's revised (1949) postulate set (39) see the section by Koch in this volume. Our analysis into  $S$ - $R$ -reinforcement terms uses the pre-1949 variant of the theory in which these problems have most commonly been discussed before.

It might be argued that this set of methodological considerations reveals a defect in  $S$ - $R$ -reinforcement theory: the several variables affecting the dependent variable are not experimentally separable. But this is the case only in this setting; each of them can be shown in other experimental settings, and, as such, may be supposed by the usual processes of induction to be operating, even though their lack of quan-

tification does not permit prediction beyond direction of sub-effects, in this more complex case.

2. *The free-exploration type.* In these studies, as in the Blodgett type, the manipulated variable is the incentive. Rats explore a maze freely, usually under conditions of increasing deprivation of food and water. The “test” condition involves feeding the now-hungry rats either following a run or directly in the goal-box. The criterion of latent learning is met if the exploration group make fewer than chance errors, or fewer errors than are made by a naive group on their first run. An experiment in which the first feeding occurs directly in the goal-box, not preceded by a run through the maze, does not give us any estimate of the effects of the free maze exploration alone. There is one such study [Buxton (5)]; this study was positive. The effects of free maze exploration uncomplicated by any feeding in the maze can be tested in the second type by noting the error scores for the first run on which the rats are fed. Where the data are reported [Daub (17), Haney (33)] it can be seen (although these authors do not point it out) that these rats, *even before reinforcement is ever encountered*, have developed dispositions to stay out of the culs during their free-exploration period. Four such studies are reported, all showing “positive” results [Daub (17), Haney (33), Karn and Porter (42), Lashley (51)]. MacCorquodale and Meehl (58) have shown that within a 15-minute period of free exploration of a multiple T-maze, rats will significantly reduce their tendencies to enter culs, and will on a subsequent day exhibit striking tendencies to choose “correctly” on a test run. In this latter experiment no food-reward, and indeed no reward of any kind, even mere return to the home cage, had been associated with “getting to the goal-box.” Looking at the animals on their test run, it would be easy to project goal-seeking into their behavior, especially if a food-reward had been somehow paired with the goal-box.

The above data serve to substantiate some of the hypotheses, discussed above, that a Hullian could use to account for the Blodgett effect. Insofar as the free maze exploration design is a variant of the

Blodgett design in which the latent period experience is free access to the maze rather than successive runs from start to end-box, some of the considerations that show the Blodgett design to be inappropriate in the role of *experimentum crucis* show the free exploration design to be similarly inappropriate.

The next two types of latent learning study differ from the above in that the rats, during the latent phase, encounter an incentive for which they are operationally satiated. They may or may not be irrelevantly motivated and reinforced, and presumably some rewards are operating even in the satiated case, e.g., removal from the restraint of the maze, return to the home cage, etc. On the “test” trial the rats are run, after a period of deprivation, for the previously encountered but unwanted incentive.

3. *Incentive present with no competing strong drive.* Rats operationally satiated for food and water are given a series of runs in a T-maze (except that Szymanski used a multiple-unit T-maze) with the incentives present in opposite arms of the maze, or in one case, with both food and water present in the same arm [MacCorquodale and Meehl (57)]. Seven positive studies using this design have been performed [Meehl and MacCorquodale (64), MacCorquodale and Meehl (57), Spence and Lippitt (100), Spence, Bergmann and Lippitt (98), Seward, Levy and Handlon (87), Szymanski (104) and Thistlethwaite (107)]. There are two negative studies [Kendler (45) and Maltzman (61)]. It is interesting to note that in both of the studies in which it was tried [Spence and Lippitt (100), Meehl and MacCorquodale (64)] the rats were, after a successful “test” under conditions of deprivation for one of the incentives, unable to shift their responses when the deprivation conditions were shifted to the alternative incentive, and, at least in the Meehl and MacCorquodale study, the first, successful test response had not been reinforced.

This design is considerably freer from methodological defects than either the Blodgett or free-exploration designs. A positive result using it (and the percentage is higher than for most other designs) is more

embarrassing to the *S-R*-reinforcement theorist. He can deduce a molar-descriptive "cognition" by positing the secondary reinforcement of fractional antedating responses to the goal-objects during the latent phase. To the extent that the drive inducing operation prior to the test run also strengthens antedating responses appropriate to the deficit, it may mediate correct turnings on the test (64). The difficulty with this explanation is the general vagueness of the  $r_G$  construct as to the conditions of its strengthening, its role as *elicitor*, and, finally, its specification as to locus. (See below for a discussion of the  $r_G$  concept in its relation to "cognition.") The implications of one such attempted derivation of descriptive cognition by use of the  $r_G$  concept (64) have been experimentally tested by Thistlethwaite (107) with results adverse to the *S-R*-reinforcement analysis.

4. *Incentive present with strong irrelevant (competing) drives.* Of eighteen such studies, seven are positive [Bendig (1), Christie (14), Dieserth and Spence (22), Strange (102), Thistlethwaite (108), Walker (182), and Walker, Knotter and DeValois (183)]; and eleven are negative [Christie (13), Fehrer (28), Gleitman (31), Grice (32), Kendler (46), Kendler and Mencher (49), Kendler and Kanner (48), Littman (54), Spence and Lippitt (101), Shaw and Waters (88) and Walker (180)]. The difficulties in accounting for the discrepancies between these studies are enormous. If one draws up a master table comparing the possibly relevant details between the positive and negative studies of this group, and with the generally positive studies of the previous (non-competing drive) group, he finds he can make only the broadest generalizations, to each of which exceptions can be found.

The only systematic difference between Types 3 and 4 seem to be the basis on which they were classified: whether the irrelevant incentive was encountered with or without a strong competing drive. In general, this design, in which the animal is tested for his cognitions of goal-objects encountered while he was strongly motivated for a different goal-object, takes too little account of other, explicitly stated, aspects of the cognition view to be a wholly adequate, or even fair,

test of it. The “emphasis” value of an incentive is small when the organism is satiated for that incentive; it is less, or even frustrating, when he is motivated for another incentive, which he has encountered in this situation before. A further, and related, difference between these is that when the organism is motivated for one incentive during training he must be forced to the side of the T-maze containing the irrelevant incentive (except in those cases where the relevant incentive is placed in both arms of the maze); forcing is suggested as conducing to negative emotional conditioning to the side of the test incentive.

We have been unable to detect any clear-cut differences in method between the positive and negative studies within Type 4, incentive learning with strong competing drive. Although the drive *present* during the latent period has usually been thirst, these are about equally divided between the positive and negative, and of two studies in which the training drive was hunger, one is positive, one negative.

There is some evidence to suggest that irrelevant incentive learning is more likely if the relevant incentive during the latent period is symmetrically located in the maze, i.e., rats motivated for water find water on one side, food *and* water on the other. The necessity of forcing the animals to the side containing the irrelevant incentive is reduced by this method, and studies in which forcing is absent or minimal (involving a small percentage of runs) tend to give positive evidence of irrelevant incentive learning more often than studies in which forcing characterizes half or more of all trials [see Walker, Knotter and DeValois (183) for direct evidence on this variable].

The attempt to favor discriminability of the two arms of the T-maze by painting them black vs. white, by using a marked illumination difference in the two sides, or by using different coarseness of screening on the floors of the two sides, seems to be ineffective if not somewhat contrary to the aim; this is apparently due to the fact that such discriminative cues conduce to position habits. Most studies in which black vs. white alleys are used show a strong avoidance response to the white alley. Walker, Knotter and DeValois (183) have shown that rats

who have developed strong position habits during the latent learning period are less likely to respond appropriately on the test trial, after motivation is shifted.

Training in drive discrimination, or at least experience in deprivation of the test incentive to match that for the training incentive, might favor irrelevant incentive learning (13). It would be profitable also to investigate the effect of providing additional alleys on the test trial, to discover whether those rats which shift performance on the test trial are *avoiding* the previously preferred alley (in which event they would choose equally among alternatives) or are indeed responding on the basis of earlier encounters with the incentive for which they are now motivated (in which case the alternative, "neutral" alleys should also be avoided).

The last latent learning design, Type 5, involves a first phase in which animals are run in or permitted to explore T-mazes whose goal-boxes are tactually or visually discriminable, usually hungry but without incentives present. The rats are subsequently fed *directly* in one goal-box, i.e., they are placed directly in the goal-box, or run through a straightway to it. The test of the learning of maze characteristics is made by reintroducing the rat to the entry box of the maze and observing his tendency to choose the side which leads to the goal-box in which he was fed. Tolman and Gleitman, after a training phase in which the rats were fed in both goal-boxes, shocked them in one end-box (withholding the usual food) and fed them (as usual) in the other—these experiences not following an actual run. The criterion measured was subsequent avoidance of the shock side of the maze. Four positive studies of this design have been reported [Tolman and Gleitman (159), Iwahara and Marx (41), Gilchrist (30), Seward (85)]; three are negative [Leeper (52), Denny and Davis (21), Seward, Datel and Levy (86—two of three replications, the third was positive)].

This design appears to be most like Type 3 in constituting, when the results are positive, an embarrassment to the *S-R*-reinforcement theorist. Here the  $r_G$  concept seems to us to be even more strained in

use; the  $r_G$ 's acquired during the latent phase are unrelated to the goal object and minimally reinforced. The "reinforcement" in the goal-box not following a run might strengthen the  $r_G$ 's peculiar to that goal-box and thus the turn leading to it on the next run. But here, again, the lack of independent verification of the  $r_G$  event and its non-quantitative status reduce its utility.

We have four general comments to make upon this extensive and rather puzzling experimental literature. First, and of greatest importance, is the necessity for sufficient replication using each of the five designs to justify confidence that a describable design does consistently yield a certain (positive or negative) result. There is no satisfactory classification of experimental setups, even using multiple criteria, which permits a clear sorting into "positive" and "negative" outcomes, although Thistlethwaite's review (106) is a valuable contribution in that direction. One can only conclude that there is something wrong with our controls or with our standards for what constitutes an adequate description of them. It cannot be too strongly emphasized that there is little point in disputing about the causal interpretation of an experimental effect when we have not yet apparently described the conditions sufficient for obtaining it.

Secondly, even if finer sub-classes of designs can be finally made replicable, one must beware of assuming that because they are all called "latent learning," they must be explainable in the same, or even a similar, way. It is perfectly possible, for example, that the Blodgett effect is based upon satiation of exploratory drive in the experimentals, and is further enhanced by the drive-conditioning mechanism we hypothesized above; and yet that the positive findings which preponderate in Types 3 and 5 will ultimately find *their* explanation in cognition-theory terms.

Thirdly, cognition theory must be developed beyond the point where its proponents' main use of latent learning studies is in the form of criticizing "the opposition." The demand "How do you explain *that* one?" is, to be sure, not wholly obscurantist, since any finding clearly

adverse to a theory is, indirectly, favorable to its competitors—provided they already have some inductive support of their own. But it seems legitimate to insist that failures to get “latent learning” pose a serious explanatory problem for cognition theorists, just as positive results do for the *S-R*-reinforcement view. Cognition theorists are not entitled to take the position that any positive design gives them their case, in spite of the negatives. Thistlethwaite has argued very insistently that “... it must be possible to demonstrate for *each* instance of latent learning or of irrelevant-incentive learning (1) that some source of reinforcement was operative in the experimental setup ...” (106, p. 120). The word *demonstrate* in this quotation needs scrutiny. In any developed science, when we attempt to understand the complex case, we recognize that there may be great difficulties in experimental separation of the determining variables. Actually, most scientific experiments presuppose (i.e., do not attempt to prove, independently) a whole network of causal laws, including those relating hypothetical constructs, that have been confirmed in other research. Even the use of complex measuring apparatus would be impossible if this sort of inductive extrapolation were forbidden. So, we can accept Thistlethwaite’s desideratum only if a rather weak meaning is given to his word “demonstrate,” somewhat like “render probable on the basis that factors present in the design have been shown in other designs to act as reinforcers.” If this weakened demand cannot be met, grave doubt is surely cast upon *S-R*-reinforcement theory. Whether such doubt gives *strong* support to cognition theory depends upon the difficulties of the latter, *and* upon the plausibility of still other causal models (e.g., *S-R*-theories not stressing reinforcement). Intermediate types, such as Seward’s (84, 85) and the theory we shall present below, must also be considered. One can imagine an accumulation of experimental evidence on latent learning alone which would render both *S-R*-reinforcement theory and cognition theory thoroughly untenable, neither being capable of generating correct predictions as to the pattern of positive and negative results.

Finally, definitive interpretation of the experiments is rendered difficult by a terrible lack of data regarding certain quantitative matters, both in theory and at the descriptive level. It is customary for Tolman's critics to attack his formulations on these grounds, but it is not sufficiently appreciated how vulnerable we all are in this respect. For example, *S-R*-reinforcement reconstructions of "positive" latent learning designs characteristically invoke  $r_G$ , and require suitable quantitative assumptions about such matters as the "similarity" of the afferent consequences produced by fractional eating and fractional drinking. It would be an interesting exercise to bring together all of the applications of this device with an eye to the question, whether the quantitative assumptions required by *S-R*-reinforcement theory to "deduce" experimental outcomes in the various settings are compatible. It is refreshing to reflect upon the disconcerting consequences of a direct experimental attack upon the hunger-thirst "stimuli," such as Heron's (36), for such theorizing. We have never seen an empirical curve for the generalization gradient of any "pure stimulus act" in the rat, nor even any rigorous theoretical treatment of it. It hardly seems quite proper for *S-R*-reinforcement theorists to attack Tolman on this front, when they themselves have frequent recourse to such an elastic explanatory construct as  $r_G$ .

In spite of the preceding difficulties of interpretation, it seems safe to say that the current state of the evidence is at least encouraging to the theorist oriented to some form of expectancy theory. We were, frankly, somewhat more impressed by the overall trend of the evidence than we had expected to be. This reaction led us to attempt a somewhat greater degree of formalization of the expectancy position, which we present in the following section.

#### IV. SOME TENTATIVE PROPOSALS AS TO FORMALIZATION

The commonest criticism of Tolman is that he has failed to make his theory "explicit" and "rigorous," the critical emphasis often being

on “quantification.” The matter of explicitness and rigor has been treated in Part II, and this criticism seems, in the main, to be valid. As for quantification, the most casual reading convinces one that Tolman either is not strongly interested in it or feels it to be premature. In a recent article he says:

Curves could be fitted. Equations for these curves could be mathematically determined and the magnitudes of the constants could be found. In fact, all the precise techniques of quantitative method could be elegantly carried out ... and bring about closure for all those psychologists who are probably at heart mere physicists or perhaps mathematicians gone wrong (151, p. 147).

And in the same paper we find the following:

And although, as usual, I have been merely programmatic and have not attempted to set up, at this date, any precise systems of postulates and deduced theorems, I *have* made some specific suggestions. ... I feel that once we have thought of really good defining experiments ... we can then hypothesize equations, fit empirical curves, and dream up constructs to our hearts' content (151, p. 154).

If we ignore the touch of sarcasm, and the suggestion that physicists are somehow more entitled to an interest in quantifying than are psychologists, what can we see in these quotations? The description of himself as “... as usual ... merely programmatic” suggests that Tolman believes in the *ideal* of a quantified science. Indeed, it would be difficult to doubt this in a writer who for many years has presented us with “functions” of “variables,” albeit in the most general form possible as “ $f_1$ ,” “ $f_3$ ,” etc.

It is perhaps worthwhile to pause a moment and consider the degrees of quantification a behavior theory may attempt. Without claiming any sort of exact scale for degrees of quantification let us present a crude ordering system for convenient reference. At the lowest level, we have (1) statements as to *which* variables determine the dependent variables. This already has an element of quantification since it claims a non-zero value for certain partial derivatives and

assigns a zero value to all others. Such a list of “relevant variables” almost always includes a further degree of quantification, (2) a statement of the sign of the first derivative. This latter may include a statement of the range over which this direction of influence holds (e.g., “monotonically increasing” if it holds throughout). Frequently, but by no means always, we are given (3) statements of signs of higher derivatives, e.g., “monotonically increasing decelerated function.” Without specifying numerical values, we may be told (4) the *order* of sizes of the various first partial derivatives. Thus, “ $x_1$  affects  $y$  more than  $x_2$  does.” It is often difficult, in the case of dimensions differing qualitatively, to say just what this latter kind of statement means even when nothing but order is claimed. Still further, we may have (5) various degrees of specification of the function *form*, as “a decay function,” “a second-degree polynomial,” and so on. Finally, we come to (6) estimates of the parameters in these functions.

No one can prophesy whether making an effort to attain one of these levels of quantification will be fruitful at any given stage of knowledge. Certain of Hull’s later efforts, both experimental and postulational, seem to indicate that he believed work at level (6) to be in order, and quite obviously he had been for some years thinking at level (5). It is important to see that there are intermediate degrees of quantification, so that a theorist may be very cautious about levels (5) and especially (6), but still may attempt more than Tolman has to date. One does not need to be “anti-quantitative” to raise the question whether the painstaking determination of parameters arising from the study of a particular species, drive, response and apparatus is not somewhat premature. On the other hand, there is a degree of quantification which is necessary for any *important* sort of “explicitness.” That is, one may set his sights lower than levels (5) or (6), but he can hardly attempt a usable (prediction-generating) set of postulates if he stops, say, at level (1) or (2). Unfortunately, Tolman seems to have done just this. It is questionable whether we can even proceed with the search for the “really good defining experiments” of which he speaks

in the second quotation, until some efforts have been made at quantification levels beyond (2).

There is no important “system” at present which does not use hypothetical constructs or incompletely reduced intervening variables, such as some form of the learning-performance distinction. When Tolman speaks of “defining experiments” for each of his “types of learning,” we have to remember that such defining experiments will in each case involve an inference *from* facts about response strength *to* values of the hypothetical entity or process. Thus, we cannot study the acquisition of an equivalence belief without having some notions about asymptotes of the cathexes being utilized in such an experiment. Again, we will be utilizing some instrumental act, so our discussion will be bound up with assumptions about the animal’s field-expectancies. If we were to interest ourselves, say, in such a curve-fitting question as: “In what way does the rate of acquisition of an equivalence-belief depend upon the cathexis of the final goal?” we would need to have *some* quantification of final goal-object cathexes for the abscissa, and it is evident that cathexis is not a word in the data language. We know the kind of experiment Tolman uses for this preliminary problem—it appears in Chapters III and IV of *Purposive Behavior in Animals and Men*. But such experiments, while they are the obviously appropriate beginning, can only establish an ordinal series of cathexes, which is insufficient for purposes of a defining experiment on equivalence-beliefs unless it is to deal with order only. The point is that *somewhere* in the system of assumptions there must appear a minimal degree of hypothesized quantification in order to get a foot in the door prior to an empirical determination via a defining experiment of any of the theoretical quantities.

This is, of course, only one of the several forms in which the same fundamental dilemma recurs in all theory construction. We are constantly in the position of leaning upon *some* poorly confirmed “laws” as we try by experimentation to tease out others. So long as we are aware of this and see clearly the mutual support aspect of the com-

ponents of a theory, there need be nothing disturbing in the situation. Such free play, with limitations, has to be expected in any ongoing inductive enterprise.

The chief aim of the Dartmouth Conference was analytic and critical rather than synthetic-constructive. However, the great stress usually laid upon lack of explicitness in the criticism of Tolman impelled us to tentative efforts toward developing his views in this respect. We cannot emphasize too strongly that what follows could most appropriately be given some such Teutonic title as "Prolegomena to an introduction to proposals for the skeleton of an incomplete postulate set generating a modified expectancy theory." In order to avoid repetitive apologizing, we shall use the unqualified terms "axiomatization" and "formalization" to designate our procedure. But we hope we are under no illusions as to the appropriateness of these designations. There are, in particular, two misinterpretations of our intentions which we wish the reader to guard against.

1. We do not propose the following as a sufficient set of "postulates" for the derivation of all or even most empirical behavior-laws. Nor have we made a formal effort to investigate them as to independence or consistency. The reader should look upon them as "programmatic," but less so than Tolman as he stands.

2. They attempt to explicate not so much "Tolman" as "an expectancy theory." While we have tried to be oriented by Tolman's emphasis and have not knowingly contradicted him, it would be absurd to expect that Tolman himself would accept all of the following as falling within the range of his own broader hunches. We will begin with a few general remarks to clarify our orienting attitudes toward an expectancy theory. As a first approximation to be elaborated as we proceed, we could express our notion of an "expectancy-type" of theory as follows: An expectancy-type of theory is a learning theory in which the fundamental learning construct, the "what-is-learned," is notationally specified by explicit reference to the stimulus-event which has commonly terminated a stimulus-response sequence

in an organism's history (the "expectandum"), in addition to making the usual reference to the eliciting stimulus and the (strengthened) response. A correlated feature, without which this notational practice would be sterile, is that the activation postulate which relates this learning construct to a behavior-disposition makes explicit reference to this third element; whereas in non-expectancy theories the "usual terminator" (e.g., reinforcer) of an *S-R* sequence plays only a historical role, in the sense that it occurs as an experimental factor in the acquisition postulates but is not "part of" the learning construct itself nor, therefore, available for use in the activation postulate.

It is our belief that a phrase such as "field-cognition" or "field-expectancy" is confusing, since it links together two words the emphases of which, while perhaps not wholly independent, are at least distinguishable (cf. Spence, 96). The term "field" makes a reference to the Gestalt and perceptual emphases in Tolman, while (in our opinion) the term "expectancy" indicates a quite different aspect of the system. It is the latter, "expectancy" aspect which we have taken as definitive of the system, since we are not persuaded that there is any very intimate connection between acceptance of the Gestalt laws of stimulus organization and acceptance of either an *S-R*-contiguity, an *S-R*-reinforcement, or a (so-called) *S-S* theory of acquisition. The problems of stimulus equivalence, and of the physical statement of configural properties on the stimulus side which will yield stimulus-equivalence or various degrees of generalization, *will recur in any form of behavior psychology*; and it is a historical accident that the leading expectancy theorist has Gestalt-emphasizing leanings. Thus, Skinner, clearly an *S-R*-reinforcement non-expectancy theorist, would find it unnatural to connect configural emphasis in any *essential* way with the matter of acquisition laws. If experimentation shows, e.g., physical triangularity of the visual stimulus to be a basis for strong generalization effects, so be it. The specification of relations among the apices of such a "physical triangle" which yield such-and-such degrees of generalization can be achieved if desired (62).

If one speaks in terms of the raw data that give rise to a science of psychology, all behavior-data theories are “*S-R*” theories, in the broad sense. The empirical variables are movements of the organism and environmental events which occur in certain temporal relations to them. That his system is an out-and-out behaviorism is one of the things on which Tolman has been most explicit. There is, to be sure, a more restricted and less innocuous sense of the phrase “*S-R* theory” which does *not* apply to Tolman, as we shall develop below. But the point here is that, at the data-language level, Tolman is the same as Guthrie, Skinner, or Hull, and he intends to be.

Since all of the theories considered by us make the learning-performance distinction in some form, the mere making of this distinction cannot differentiate an expectancy theory from the others. Once the organism has acquired *whatever* sort of bond or knowledge the theories respectively specify, all of them allow for a manipulation of overt behavior strength by means of control of drive, fatigue, emotion, or competing responses. This means that having learned, and not having yet unlearned, an animal may still fail to exhibit the behavior. Guthrie, Hull, and Skinner all make their various places for this fact. If it was ever an adequate rendition of the Tolman position’s uniqueness to say that “The organism learns certain things but may exhibit this learning in different ways depending upon ...,” it is no longer. The distinguishing feature of an expectancy approach is not as simple as this, nor is it easy to state in any short characterization.

## V. EXCURSUS: THE RESPONSE CONCEPT

Since Tolman’s system is a behaviorism, its dependent variable necessarily reduces to some aspects of the organism’s activity. Therefore, the problem of “defining the unit of response” exists for Tolman as it does for all behavior theorists. From *Purposive Behavior* through the series of experiments on “spatial learning” (78, 79, 80, 81, 160, 173, 174, 175, 176) to his recent theoretical papers (150,151), Tolman

has obviously felt that the definition of the unit, the specification of “what is learned,” was one of the major differentiators of his position. In an earlier paper, we stated our own opinion as follows:

We early concluded that certain views which have been linked historically to Tolman’s formulation are logically unrelated to the “core” concepts of an expectancy theory. Other views seemed related to Tolman’s formulation until preliminary efforts to formalize them indicated their independent status as well. We shall merely list those dogmatically as “properties *not* definitive of an expectancy theory”:

1. Gestalt-configural stress.
2. Perceptual field stress.
3. Pure contiguity as a sufficient condition.
4. Specification of reaction-class by reference to position, direction, or locomotion (rather than by effector-properties).
5. Discontinuity view of discrimination learning.
6. Insistence upon the learning-performance distinction (66, p. 230).

In a personal communication, Professor Tolman states his agreement with us that four of these emphases are not crucial to his position, but he feels that two of them, the “Gestalt-configural stress” and the “specification of the reaction-class” are basic points on which he differs from *S-R* theorists. Consequently, we feel it necessary to digress for a somewhat more extended treatment of the response question before proceeding with our own formalization proposals.

To begin with, in denying the cruciality of the response-definition as a mark of expectancy theory, we do not mean to suggest that there is no problem here, or that Tolman tries to solve it no differently from others. But we do think it incorrect for a Tolmanite to assume that *S-R* theory necessarily means an *R*-term specified by muscle-twitches, and that experiments indicating the utility of a more “molar” *R*-term are, *prima facie*, embarrassing to non-expectancy theories. The distinction is not that simple, witness the fact that one major *S-R*-reinforcement theorist (Skinner, 90) wrote a whole book in which the reports of raw data tell us, strictly speaking, *nothing* about the actual activity of

effectors. A glance at the experimental reports of Hullian and neo-Hullian workers shows the same thing. Levers are pressed and maze “choices” are made, and the observations are narrated in these terms, for the contemporary *S-R* theorist operates with a “molar” (in the sense of levels: 55) definition of response and in practice worries very little about this aspect of his methodology. It is interesting to note that Hull (38, 40) devotes no discussion to the question of what are “admissible” modes of specifying *R*. The Glossary of his posthumous book defines response thus: “*R* = response; an act of some kind” (40, p. 358). However, it seems unlikely that all formulable *R*-terms would be equally acceptable to Hullians (e.g., the “act of some kind” being strengthened is “*getting to the goal-box*”). One meets clinical psychologists who are suspicious of the learning-theory approach to psychotherapy because they perceive *S-R* formulations as “atomistic,” as attempting to analyze personality into “discrete responses” rather than treating the “whole person.” These obscurantist attitudes spring at least in part from the inadequate attention usually given to the methodological questions devolving about the response concept.

Any effort to clarify the response problem must begin by explicit recognition that, for all learning theorists and all experimenters, the “response” is a *class*. This follows directly from a definition and a fact. Learning is defined as a change in an organism which is reflected by a change in the strength of some specified way of acting; and it is a fact that no two (numerically distinct) intervals in the flux of an organism’s activity are absolutely identical in all of their properties. Hence, any experimenter who talks about learning’s having taken place must be classifying two or more occurrences as “instances” of this way of acting, in spite of their detectable differences. This is simple and obvious, but it is one of those simple things which is easily forgotten and very clarifying when remembered. As an example of the resolving power of this class-concept, consider the objections one sometimes hears that a bit of *S-R* analysis “abstracts from the total activity,” and hence somehow distorts in describing. The obvious reply is that *all*

description which names an activity and identifies its recurrence is already abstracting. Two “instances” of any kind of behavior, classified at any level of molarity, are unavoidably abstractions, in the sense that they select certain features of the flux for attention and ignore others. So when a critic voices this sort of objection to a given formulation, he must be made to see that all conceivable formulations would be subject to the same complaint, and that a claim to be telling *nothing but* the truth about an event is never a claim to be telling the *whole* truth about that event. The individual, dated behavior-event is “unique”; the very concept of learning presupposes that we are willing to group these dated events under one class-name by virtue of certain common properties and neglect those other properties or quantitative variations with respect to which the class-members differ among themselves.

This leads directly to another important point in connection with the question of “what is learned.” It is evident that a given interval of the flux can be classified in more than one way, and that conjunctions and disjunctions of classes can be set up for study. This shows that, at the purely *descriptive* level, there may be alternative and seemingly quite different ways of narrating the flux, *all equally* “valid.” Some implications of this for the learning controversy will be developed in the course of our discussion. Tolman emphasized in *Purposive Behavior in Animals and Men* that his concentration on the lawful features of goal-seeking behavior “at its own level” was perfectly compatible with the usual assumption that the efficient causes of these behaviors lie in the physiology. We must not assume that there are no intermediate levels of analysis between these two extremes, which would exhibit a lawfulness of their own. Whether there is always some one “optimal” degree of specification of the response properties which will generate maximum orderliness as suggested by Skinner (90, p. 38) is a complex problem which cannot be discussed here. But it is an obvious mathematical possibility to obtain a very smooth curve of responding when the behavior is classified at a more “molar” level, this curve

being a resultant of the several curves which would be generated by counting responses further subclassified at a *less* “molar” level. These latter curves, perhaps equally “smooth” as the original, might individually share its form or instead exhibit other orderly (e.g., cyclical) properties. The idea of smoothness or orderliness invoked by Skinner in this connection seems to require some clarification before it can be rigorously applied to the problem of defining the response.

We begin with the complete stream of activity as it runs off in the time sequence, the rich, raw, unclassified *flux* of behavior. Any arbitrary *interval* of the flux can be demarcated by time-points and examined for the presence of certain properties. The operational specification of a *descriptive property* of an interval permits the use of words referring to the animal’s visible anatomy, the descriptive words of the physical thing-language, and special words which we define explicitly in terms of these. Observation-sentences are formulated in these words, and must not involve even implicit reference to any other intervals of the flux. There does not seem to be any cogent objection to the inclusion of words referring to *contact* with environmental objects (“The rat put his paws on the lever”), or to *orientation* (“The rat turned his head toward the light”), or to *locomotion* (“The rat approached the tray”), so long as these environmental objects and the organism’s relation to them are describable by explicitly defined words or phrases referring solely to the interval under consideration. If a rat is oriented so that his nose is pointing to a lighted circular disk, “locomotion forward in a straight line” is a legitimate descriptive property of the ensuing ten seconds of the flux, and such a specification of *R* would presumably be just as available to Spence as to Ritchie.

It is perhaps noteworthy that more “complex” flux-properties are introducible without going beyond pure description. For example, the “if ... then” (which is the “not ... unless”) relation in the sense of material implication (105, pp. 23-28) is definable in terms of the (intuitively simpler) logical constants “not” and “or” or “not” and “and.” Hence,

complex properties such as “Lifting the paw, whenever he has in the preceding five seconds wiggled the whiskers, provided that ...” etc., are clearly still descriptive. Sequence specifications, with or without a metric time-property, are also admissible.

On the other hand, non-extensional connectives such as “in order that ...” or “necessitates that ...” are clearly forbidden as going beyond the descriptive properties of the interval. If they occur, they must occur as theoretical discourse, not as pure description.

In practice, we do not begin by slicing up the flux into a huge (theoretically infinite) set of intervals and then by examining each one for the presence of the desired property. There are certain “striking” features of the flux which, in addition to our theories and our anthropomorphic identifications, help us direct our attention to some intervals and, for that matter, to certain properties. That theory and anthropomorphism are among the *causal* sources of our experimental attention is of course quite irrelevant to the question as to whether what we emerge with in our narration is itself purely descriptive.

Some of the more obvious features of the flux which make our trial-and-error not arbitrary but rapidly convergent on the relevant properties and intervals may be briefly mentioned. Suppose a non-anthropomorphic and non-theoretical experimenter blindly plotted time-lines of numerous distances and angles in the flux (e.g., angle at elbow-joint, distance of chin from floor). How would he “notice” promising regions of the flux, avoiding the trying out of tentative *R*-classes which would be theoretically and predictively useless?

1. *Inception and termination of contact with an environmental object*: The rat touches the lever and his pressing of it is usually followed by a breaking of this contact.
2. *“Phasic” property*: Most extreme excursions of any effector-position measure are followed shortly by an undoing. Rarely does rat or human lift the head high or greatly reduce the angle at a joint without shortly reversing the change. The blindly plotted time-lines would thus exhibit a tendency for cycles and

- overshootings in returning from an extreme excursion.
3. *Segmental*: An obvious possibility is to examine all intervals in which the flux shows changes of a considerable magnitude in the measures on a given limb or other anatomical region.
  4. *Multimodal distribution*: Distribution of excursions may be multimodal. Thus, the distance of a rat's nose from the ceiling can take on all intermediate values within a certain range. But the plot of this distance for numerous intervals would show a small mode near the upper extreme—these being chiefly occurrences of the exploratory reaction "standing up, sniffing ceiling."
  5. *Intrabehavioral covariation*: Families of the arbitrary time-lines would show concurrent variations of activity; e.g., "speech" involves a burst of excursions and cycles of reference points and lines on lips, jaws, tongue, diaphragm, glottis. Such a highly configured pattern as a "shrug" might be isolated and recognizable after only one occurrence of it.
  6. *"Achievement"*: This very common, if not commonest, criterion raises such difficult problems that it will be separately treated below.

Having found an interval  $I$  in the flux which is identified by its possession of a simple or complex property  $P$ , what is "the reaction" ( $= R$ )? It does not consist of the whole interval, because we want to be able to speak at times of several "concurrent reactions" all of which take place *in* the interval or an interval  $I'$  which overlaps  $I$ . If the property  $P$  includes a specification of a certain body segment or effector-system, the reaction  $R$  may be defined as the movement of the relevant part(s). If  $P$  has been defined as a disjunction of alternative parts or movements, the reaction consists of those which verify the occurrence of  $P$  in the given instance. Thus, if "touching lever with paw or paws" is  $P$ , the reaction  $R$  is the touching by right paw when that happens, but if instead both paws had been used, the movement and touching by both would constitute the reaction. Note that we

cannot speak of *instances* of the reaction *R*. The reaction *R* is dated, a unique, unrepeatable *event* in the flux. It has “parts” but it has no “instances.” Any parts of the reaction are in the relation of part-to-whole to it, and link-in-chain to one another; but not in the relation of class-membership to it. The *R-class* has instances, each of which is a reaction. Elsewhere herein we shall use the term “response” as a synonym for “*R-class*,” the family of reactions identifiable by their common possession of some property *P*.

We have now to consider the common procedure of defining an *R-class* in terms of *achievement*. This practice ranges from the usually non-controversial case in which the achievement is a simple, concrete manipulandum-event (“Rat pressed the lever”) to very complex and debatable instances such as occur in describing human social behavior. (“He kept at it until he had put Smith in his place!”) It is in this type of *R-class* specification that the line between pure description and covert theorizing becomes difficult to draw. It is here also that the lines of battle are likely to be drawn between Tolmanites and others insofar as the response becomes an issue between them.

First of all, one must distinguish between what the experimenter bothers to narrate (or is ingenious enough to discern) and what he can be confident is objectively the case. In point of fact, any *R-class* specified by achievement is a truth-function of a finite set of *R-classes* specified by non-achievement flux-descriptions. There is more than one way to get a lever down, but there is not an infinite number of ways. (A “way” here is a sub-class defined by fairly restricted effector topography, but necessarily allowing for certain quantitative variations over its own instances.) Whether we as experimenters have the skill or make the effort to state what they are, there is some finite disjunction of effector-event-classes matched to any manipulation- (or locomotion-) attainment class. Presumably, this is why workers of varying theoretical persuasions seem equally comfortable with describing the two most commonly studied types of rat-behavior in achievement language. The data emerging from Skinner-box work are obviously of

that sort. As for the maze, the word "turn" may perhaps be construed as an effector-event word. But the use of "choice" as a common synonym, or electrical recording or opal-flash glass where we cannot see the rat, suggests that "turn" is, like "press," ordinarily an achievement-word.

The most hard-headed operationist can afford to be relaxed about this use of achievement-words in description because studying the quantitative laws of such achievement-classes need not (in fact cannot) commit him immediately to any particular theory about the learning process. To the extent that docility with respect to a goal or sub-goal is *really* immanent in behavior, *really* part of behavior's "immediate descriptive warp and woof" (128, p. 12), all theoretical issues remain to be settled after such behavioral descriptions have been given. It is important to see that the lawfulness exhibited by attainment-specified *R*-classes does *not* imply, or even suggest, that the coordinated flux-disjuncts are not lawful. Nor does a finding that attainments show lawful changes tend to show that attainments, *rather than* the coordinated flux-disjuncts, are "what is really being learned." From the fact that a rat presses the lever now with left paw, now with right, sometimes with teeth, we cannot infer anything about "what is learned," but must undertake further behavioral analysis to tease out the conditions for occurrence of each of these topographies. After all, if a rat "learns to press with chin" and "learns to press with paw," it is a logically trivial consequence of these statements that he "learns to press."

An achievement-learning *may* be interpreted as more than purely descriptive, of course. This ambiguity is impossible if the achievement-statement is applied to a single interval, since "At 12:07 the rat depressed the lever to the point that an electrical contact was made" is pure description. It is not in terms of effectors, and it is very incomplete ("abstracted") as a characterization of the interval. But it is pure description nonetheless. However, if I say "Over the hour the rat learned to press the lever," it is not clear how much I intend. I may

still intend pure description, which expands to something like “Over the hour there was a significant increase in the rate of occurrence of effector-events which had the mechanical effect of depressing the lever.” This amounts to “Over the hour there was a significant increase in the rate of occurrence of [foot *or* teeth *or* jaw] contacts which ...” This is all quite innocent of theory, and the variations are matters of dividing the classes into finer classes, depending upon one’s desire for detail in narration.

The interesting question is, what else can the original assertion be taken to mean, which would *not* be purely descriptive? The alternatives are not so easy to list, but they vary from a slight inductive extrapolation, such as “If the rate of  $R_1$ -disjunct declines, some other one  $R_x$  of the whole family of effector-event-classes coordinated to the attainment-class  $R_L$  will increase in strength enough to maintain the  $R_L$  level (Cf. 91, pp. 211-212)” to a layman’s anthropomorphism that the rat “knows the idea is to get the lever down and would tell us as much if he could talk.” Even the simple inductive extrapolation about a compensatory increase in one effector-event-class is often used as a counterfactual conditional, “If we *had* prevented  $R_i$ , then  $R_j$  *would have* been employed to do the job.” This asserts a lawful relation not immediately established by the flux and is not a truth-function of the observation-sentences.

This is one disadvantage of the achievement type of description: it is capable of being over-interpreted. However, ambiguous achievement-statements can always be explicated upon demand, in order to see whether their maker does intend more than description. Why, then, are there any arguments about this? Leaving aside some real disagreements as to the reproducibility of certain factual outcomes, there may be differing inductive extrapolations on the near-descriptive level, and differing theoretical interpretations based upon these.

In a recent methodological paper (47), Kendler rejects the question “what is learned” as being an unsuitable question for the learning theorist. We would prefer to take as a provisional hypothesis that

arguments over “what is learned” *are* about something, and try to tease out what this something is. We would see “what is learned” as what Carnap calls an explicandum (8, pp. 513, 517-521, 531-532), the task being the formulation of some more precise explicata as alternatives. There is no cut-and-dried way to do this; but for any proffered explication of the initially vague question we try to see how the experiments usually cited in controversy and how the words used in associated discussion would fit. In the end we will have formulated somewhat more exact questions even if they are rejected as not being “adequate” explications by the original protagonists.

Proceeding thus, we may ask, “What sorts of causal analysis of an achievement-statement would be rejected by those theorists who like to stress that the rat ‘Learns to press the lever’?” Consider the classical experiments of Muenzinger (72, 73) on the lever-pressing behavior of the guinea pig, which Tolman cites (128, p. 171; cf. also p. 18) as evidence for behavior’s “multiple trackness”: the use of alternative paths to attain a given end (achievement). Muenzinger studied the variations in the effector-pattern of a simple lever-pressing response through 1000 reinforced occurrences, long after there was no further possibility of “improvement” insofar as latency, speed, or elimination of competing responses were concerned. He classified the pressings into nine topographies—three emphasizing left foot, three right, and one each with both feet, teeth, and “head” (lower jaw). Working thus at the asymptote of the achievement-class, he found what Tolman views as substitutability of means—a rise and decline of strength among the competing effector-classes, this competition persisting throughout the course of the experiment. Muenzinger himself believed that the habit should have become “mechanized” in the sense that one specific effector-pattern should steadily increase until all others are excluded, according to the law of effect; and he argued for an interpretation in terms of the “meanings” assumed by environmental objects. But how much does such an experiment actually show? Examination of Muenzinger’s tables shows that the subclasses themselves are by no

means chaotic in their temporal changes, but exhibit a certain lawfulness. What is to prevent the *S-R* theorist from saying that, after hundreds of trials, *all* of the topographies were at fairly high strength, so that whenever one fell momentarily behind there was always a competing one available to “replace” it? (Cf. 25, 91.) The point is that all nine effector-disjuncts *did* get the lever down, hence were all repeatedly reinforced. At the time Muenzinger wrote, the laws of frequency and effect were stated more crudely than they are in any current *S-R*-reinforcement theory, so Muenzinger may have been quite right in supposing that his experiment was adverse to them. But does a more sophisticated and elaborate effect-and-frequency theory have any obvious trouble here? Suppose we view the “lever-pressing” as the environmental outcome mechanically coordinated to a set of nine effector-event-classes. Members of any of these nine “succeed,” i.e., are reinforced. If  $R_1$  and  $R_2$  designate two of these classes, does *S-R*-reinforcement theory imply that the “initially stronger” must come to dominate to the exclusion of the others? Consider just a few of the possible complications which interfere with any such easy inference:

1.  $R_1$  may be stronger than  $R_2$  initially but may have a lower asymptote.
2.  $R_1$  and  $R_2$  may have different growth rates. Nor is there any assurance that higher growth rates will be associated with higher asymptotes, and of course “accidents of history” might reverse the order of either or both of the above in relation to initial strength.
3. The evidence (92) indicates that some *R*-classes are “harder,” i.e., mechanically difficult, requiring more work, or are even painful, so that their strength may decline faster either with or without reinforcement. So  $R_1$  might have an initial edge in strength but fatigue sooner than  $R_2$ .
4. The number of responses put out per reinforcement (“extinction ratio”) varies from  $R_1$  to  $R_2$ . Hence, a given reinforcement-probability for an  $R_1$  which is not 100 per cent mechanically

effective may maintain it at a stable rate which is lower than that of an  $R_2$  about equal in effectiveness.

5. Since mechanical or emotional prevention of an  $R_1$  results in a "bottling up" effect, so that the momentary rate when the constraint is removed is much in excess of the previous rate (24) (27) (43, p. 71) (90, pp. 345-350), it may be that a similar effect occurs when a run of competing instrumental responses  $R_2$  prevents emission. In that case any of numerous factors capable of transitorily decreasing the competitor may lead to a burst of  $R_1$ .
6. Extinction is accompanied by irregularities, which Skinner considers chiefly "emotional" in character.
7.  $R_1$  may be effectively on a lower schedule of reinforcement than  $R_2$ , giving it less strength but in the extreme case greater resistance to extinction. If later in the series fatigue or a decline in drive begins to alter the  $R$ -forms slightly so that their mechanical efficacy is lessened, the weaker may catch up again.
8. Spontaneous recovery occurs and the parameters are not the same for all  $R$ 's.
9. It is not known whether induction between responses is quantitatively symmetrical. Possibly the increment in strength given to  $R_1$  due to a reinforcement of  $R_2$  is not the same as that in the other direction.
10. In the earlier stages of learning, the increment in strength given by a reinforcement to an  $R$  of the less probable class will exceed that given to an  $R$  of the more probable class (other things being equal) due to the negative acceleration of the acquisition function, the ordinal number of a given reinforcement being higher for the initially more probable  $R$ .

The interaction of all these factors, with suitable values of the numerous parameters involved, could lead to a very long series of cyclical fluctuations in response dominance of the sort Muenzinger describes. What looks in achievement-terms as one homogeneous mass of interchangeable "means-to-ends" may in fact be the resultant

of a very complex series of changes composed of fatigue, rest, extinction, spontaneous recovery, reconditioning, emotional upset, stimulus discrimination, decline in drive, induction (43), response differentiation, and the like. But after a long reinforcement history for a fairly "easy" response, there are always enough effector-classes at high strength to keep the achievement- $R$  stable, the whole system being restricted only by the mechanical constraints of the apparatus and the drive-level of the animal.

Where would an experimenter preferring the achievement-language stand on such an analysis? If he finds it acceptable, it indicates that his use of the achievement-language was purely descriptive; if he rejects it, we assume his use of the achievement-language springs from a theoretical preference. The proposed analysis seems to require no learning construct referring notationally to the "achievement" as such—no intervening variable tied somehow to the situation "lever-down." Its learning-theoretical elements refer only to the several effector-classes, and the relation of the disjunction of these *to* "getting the lever down" is merely a matter of causal laws in the science of mechanics, requiring no behavioral counterpart or representative.

Consider another case. Suppose a rat has always used his teeth, never his paws, in the course of an experiment. We muzzle him, find he presses promptly and at a high rate with his paws. Even if we give no further reinforcement, the paw-pressing continues to develop a nice extinction curve. Under certain conditions, this would not be of any special relevance in the "what-is-learned" controversy. For example, if this rat has been accustomed in the living-cage to operate a food-dispenser by means of a lever and is known to have used teeth and paws in that setting, the  $S$ - $R$  theorist would argue that  $R_{\text{paw}}$  to the experimental lever gets induced and generalized strength from  $R'_{\text{paw}}$  to the home-cage dispenser, and is emitted as soon as the stronger member of the experimental hierarchy ( $R_{\text{teeth}}$ ) is prevented from competing with it. In general, an observed "multiple trackness," "substitutability," or "interchangeability of means" is not theoretically excit-

ing when it can be shown that the substituted  $R$  is (1) already in the repertory and (2) under control of a discriminative stimulus sufficiently similar to the present stimulation to allow for easy stimulus generalization. "Sufficiently similar" is necessarily vague, but the principle seems clear.

A second uninteresting case is response induction. If we know for the species that whenever  $R_1$  is strong  $R_2$  is strong also (unless it has been differentiated), the occurrence of  $R_2$  as a surrogate for  $R_1$  in attainment-mediation is not likely to be invoked as an argument against an  $S$ - $R$  interpretation.

What other cases can arise? It might seem at first impossible to imagine any, for if the stimulus field resembles previously experienced fields which have acquired response control, the  $S$ - $R$  theorist can invoke generalization; and if it is so unlike the past to be "wholly new," does any theorist predict substitution will be exhibited? Let us look more closely at the lever-pressing situation alluded to above.

$S_h$  : Home cage stimulus field (cup, dispensing bar, etc.)

$S_1$  : Experimental stimulus field (tray, lever)

$R_p$  : Operating dispenser with paws

$R_t$  : Operating dispenser with teeth

$R'_p$  : Pressing experimental lever with paws

$R'_t$  : Pressing experimental lever with teeth

Assume that  $S_h \cdot R_p$  and  $S_1 \cdot R_t$  have never become strong in the home cage, because they were rarely or never reinforced with food. Both, however, have nevertheless *occurred* a large number of times in the pre-experimental history. The pre-reinforcement strengths of  $S_1 \cdot R'_p$  and  $S_1 \cdot R'_t$  in the experimental box are also very low. Now we strengthen  $S_1 \cdot R'_p$  by reinforcing in the box. Suppose that  $S_1 \cdot R'_t$  does not occur in the box. This latter is an ambiguous fact, since we cannot tell whether its failure to occur (once we have begun reinforcing "lever pressing") is due to low strength or to the ascendancy of the reinforced competitor  $S_1 \cdot R'_p$ . But suppose we now prevent  $R'_p$ . If  $R'_t$

occurs more than for a suitable control, yielding a large extinction curve, how do we “explain” it? We cannot invoke stimulus generalization from  $S_h \cdot R_t$  to  $S_1 \cdot R'_t$ , because we know that  $S_h \cdot R_t$  is weak and if generalized to the box as  $S_1 \cdot R'_t$  it should be even weaker. On the response side, can we simply invoke induction from  $S_1 \cdot R'_p$  to  $S_1 \cdot R'_t$ ? In a crude operational sense of “induction,” this *is* the answer. If the rat behaves as described, we have “by definition” succeeded in inducing strength in  $S_1 \cdot R'_t$  by first strengthening  $S_1 \cdot R'_p$ . What is there to disagree about?

Here again, as in the case of the statement “He learns to press the lever,” we have to decide between response induction as a *descriptive* or as an *explanatory* concept. As description, there could be no argument if the facts were as we have imagined. But, of course, this broad descriptive use of the concept *induction* says almost nothing. Any experiment—“insight,” problem-solving, concept-formation, maze-learning, complex verbal task—in which one response topography is strengthened indirectly by virtue of an operation we have performed on another, may in such a usage be dismissed as a case of “response induction.” All cases are covered in a rather empty and useless sense, and such a use would not be theoretically powerful because we cannot predict or control via it unless some rules for conditions which will yield such induction are stated. If a rat behaved very “rationally and insightfully” in choosing a new route when it was made available, any theorist could shrug off the problem thus posed as merely a case of induction! It seems obvious that we ordinarily have some narrower meaning of the term in mind. To help us get a lead on this narrower meaning, let us consider a series of cases of such indirect response strengthening, asking which ones would make Tolmanites happy and not, *prima facie*, gratify Hullians?

- a. Soft pressing  $\rightarrow$  Harder pressing. Once the class-character of all *R*-terms is clear to disputants, this case loses its theoretical interest.
- b. Pressing down  $\rightarrow$  Pressing sidewise. This one is a little more in-

- teresting, but still would not cause much concern in the *S-R* camp.
- c. Pressing with right paw → Pressing with left paw. This case is more doubtful still.
  - d. Paw press → Teeth press. We suspect most psychologists would rule this out as not a fair case of response induction. Why?

Inspection of this series suggests that it is mostly a matter of topography. We feel most comfortable speaking of induction when the related *R*-classes involve the same effectors and a fairly “similar” pattern of use of them, differing mainly in matters of force, distance, speed, duration, etc. It is as if we have in mind some *primary induction*, comparable to primary generalization on the stimulus side, and based fairly directly upon the biology of the species. Without an extreme degree of experimental control and a very detailed, continuous record of the entire history of the organism, such a primary induction would be hard to separate from any more complex or “high-level” basis for an indirect strengthening effect. In principle, if we know that  $R_2$  has never occurred, and find that when we strengthen  $R_1$ ,  $R_2$  receives strength and can maintain itself for a time without further reinforcement, we would infer a primary response induction. In practice, psychologists examine topography as a substitute for any such detailed record of the history, apparently assuming that if the topographies are too dissimilar, the primary induction would be negligible.

Now if primary induction could be claimed from  $R_p$  to  $R_t$ , the hypothetical experiment would again lose most of its theoretical interest. A Tolmanite might prefer to think of the process in achievement terms, but the more “atomistic” analysis is no longer immediately threatened by the mere fact of substitutability of teeth for paws. Suppose, however, that we know enough about the species to be confident that primary induction, at least to the quantitative extent required, does *not* hold between  $R_t$  and  $R_p$ . What, then, are the theoretical possibilities? One may flirt with the idea of *S-R* theory’s adding a new postulate, dealing with “secondary response induction.” It might take some such

form as the following: "If  $S_1R_1$  and  $S_1R_2$  are both  $> 0$  (on the basis of separate histories of reinforcement), the strengthening of a new habit  $S_2R_1$  will induce strength in  $S_2R_2$ , beyond any primary induction obtaining between  $R_1$  and  $R_2$  as related topographies." I.e., two responses acquire an inductive *relation* whenever they have been conditioned to a common stimulus. This strikes us as a rather odd sort of postulate for an *S-R* theory, but oddity is hardly an objection to it. The issue between an *S-R* theory thus augmented and an expectancy theory would then not be decidable on the basis of experiments concerning "multiple trackness" such as the one described. If it turned out that *other* types of experiments supported an expectancy framework, perhaps the job done by this special postulate in dealing with multiple trackness could be done by a derivable *theorem* in expectancy theory, in which case the latter theory would, *ceteris paribus*, be preferred. Whether multiple trackness can be derived without special auxiliary assumptions from the type of expectancy theory here proposed must be considered after we have presented the postulates.

A special use of the achievement language occurs in the description of certain "choice"-problem experiments. The behavior of Nissen's chimpanzees (74) seems to him best characterized as "approaching the white stimulus card," where the approach is not a locomotion but a "selection" of the white card in preference to the black. During the training period, there are two (compatible!) ways of describing the behavior changes. One can say that the response-disposition  $S_{WB} \rightarrow R_{\text{Left}}$  and  $S_{BW} \rightarrow R_{\text{Right}}$  are increasing in strength; he can also say that the response  $R_W$  ("selecting white card") is increasing in strength while  $R_B$  ("selecting black card") is not. Nissen shows that excellent transfer occurs when the spatial arrangement of the presented stimuli is up-down rather than left-right. If we have chosen the first sort of description, the prediction of the experimental result would seem to involve problems similar to those raised in the preceding discussion of multiple trackness in lever-pressing. The second type of description is free of these difficulties, but it requires *an explicit reference to the*

*stimulus side in its characterization of the response.* It is, therefore, a clear case of the achievement language, and one which does not seem so easily reformulated in non-achievement language as is the case of gross locomotion toward a light. Some of the theoretical difficulties likely to be encountered in such a reformulation are treated in Nissen's thought-provoking article.

We conclude this lengthy but unavoidable digression on the response concept by emphasizing the complexity of the problems and the fact that some of them still await analytic or experimental solution. For our own purposes in what follows, we shall use the symbol *R* in a fairly broad sense. Locomotion, orientation, and the kind of "achievement" class that can be specified by wholly descriptive language (including language describing manipulandum-events) and which refers wholly to the currently occurring behavior, are admissible sorts of *R*-class. Language which refers even implicitly to the properties of other intervals or to stimulation *not present to the organism at the time the response is being emitted*, is excluded. We do not believe that these rules are either so narrow or so wide as to prejudice any significant theoretical issue.

The reader who is accustomed to having his non-expectancy theories served up with such verbal adornments as "mechanical," "blind," "helter-skelter," or "meaningless" will be disappointed. Making psychology "like physics" is very probably an overly ambitious aim, but one easily available way for us to work at it is to dispense with these obfuscators. Thus, if we want to distinguish in what follows between an *expectancy* and a *stimulus-response connection*, we shall not find it necessary to label the latter a "*mere stimulus-response*" connection. We would not take Tolman's approach as a serious scientific one unless we assumed that its spirit could be caught by the postulate proposals themselves, without bolstering it with any such verbalisms.

We have concluded that the usual distinction "*S-S*" versus "*S-R*" should be sought in the more complex relations *among* constructs, i.e.,

their role in the whole system of laws, rather than in any differences in how the empirical variables are defined. Thus, we shall argue that the concept *expectancy* is more conveniently introduced by implicit definition than by any such direct “operational” reduction-sentence as that employed by Tolman, Ritchie, and Kalish (1937). Even if the supposed “operational definition” were an acceptable one, which we have argued above it is not, it does not seem to us a step in the fruitful direction. Is there nothing in common between an “expectancy” in the spatial learning problem and an “expectancy” in the Skinner box? If not, this definition of expectancy would be tied to a certain design, in a way which can hardly be the intention of a general theory. Of course, another reduction-sentence can be written for the Skinner box, and the same word used. But is this really what Tolman has in mind? Presumably some *general laws about* expectancies are envisaged. Why not, then, make the concept general from the first?

The history of experimentation, particularly in the latent learning area discussed in the section preceding, strongly suggests that it is unrealistic to search for an *experimentum cruris* which will strictly force a decision between an expectancy theory and others. Probably many experimenters designing latent learning studies have had some “crucial” intentions, but with over 49 such studies at the time of this writing, the controversy continues and new experimental designs continue to appear. It seems that we can only hope to proceed by partial formulations which entail, in an “other-things-equal” sense, certain rather broadly specified consequences, and the gradual accumulation of confirmed or disconfirmed consequences will progressively support or disconfirm the theory. Whether or not one agrees with most of his arguments, Thistlethwaite’s review (1966) points up the complex character of our inductive problem. We are in relative ignorance of many factual matters, of which the following seem to us among the more important: hunger-thirst interaction, the arousers and satiators of the “exploratory” drive, alternation effects, emotional consequences of frustration, dependence of “behavioral oscillation” parameters upon

overall drive level, the safest methods of guaranteeing “satiation,” effects of past experience upon the goal-character of objects, relative value of primary and secondary reinforcers as rewards, potency and discriminability of the proprioceptive input (particularly as it arises from damped consummatory responses) as a source of either reinforcement or discriminative control, and so on and on. New experimental data on any one of these may necessitate large revisions of our thinking about whole families of experiments. A possible example would be Christie’s work on pre-experimental experiences with the incentives (12, 13, 14). *No available theory*, it seems to us, says enough “in advance” about *all* of these and similar questions, to rule out all possibility of recourse to various *ad hoc* explanations after the fact. To be sure, any such auxiliary hypotheses must themselves have further factual implications if they are not to be rejected as *purely ad hoc*. Even so, when we set up experiments to test the implications of these special hypotheses, we find the same problem recurring afresh. No one can be “blamed” for this, for the interdependence of theoretical constructs is in the nature of science. Over 20 years ago such logical empiricists as Schlick were pointing out that there is *a sense* in which any single experiment, even though conceived with reference to a specific hypothesis, can be regarded as an indirect test of the entire causal framework within which the scientist operates. The physicist “takes for granted” the accurate working of his ammeters while he is engaged in studying something other than instrumentation problems; but a complete explication of how a pointer-reading is *able* to confirm a statement concerning mesons would, of course, have to “unpack” the internal physics of the ammeter. Hence, his successful prediction of the experimental outcome furnishes a small increment to the inductive support for the latter. One difficulty with contemporary learning theory is that our ammeter-reading is an animal’s “response-strength”; the behavioral analogue to the physics of the ammeter (or, better, of the whole experimental circuit) would, strictly speaking, have to include the answers to all the questions in the above list; and in the

present state of our factual knowledge it simply does not.

It is easy, for example, to write an “operational reduction” of, say, “expectancy.” We may take a Carnap reduction pair (7) as our model, and treat “this rat is hungry” analogously to “this sugar-lump is soluble.” But this is too easy. Carnap could get by with it because (1) it takes very little to specify the conditions, and (2) the specified conditions are readily realizable. When Carnap says “If put in water, then, ‘It is soluble’ is equivalent to ‘It will dissolve,’” no one needs to be told that, e.g., the solvent used as a test-substance must not already be a super-saturated sugar-solution! This kind of example presupposes a whole set of “obvious” and “unstated” test-conditions. But the point is that they *can* be stated, and *their own* test-conditions given by additional reductions, upon request. We, unfortunately, are not in comparable circumstances. In order to make “This rat is expecting food at *L*” equivalent to “He locomotes to *L*,” we have to include such statements as “This rat is not frightened” and “This rat is not angry” and “This rat’s exploratory-drive with respect to *M* and *N* is near-zero” *among the realizable test-conditions*. To the extent that the law-structure defining *them* is incompletely worked out, this task cannot be immediately carried through. Of course, one can take a super-conventionalist line and flatly claim that he offers the reduction-pair as his introduction of the notion “expectancy.” No one can quarrel with such a definition, since (at least if it is a bilateral reduction-sentence) it makes no truth-claim. But within the context of discovery, we like to know that a proposed formalization catches the *intent* of the initially vague explicandum, because as a matter of scientific history, if it does not, it may have lost in conceptual fruitfulness what it gained in precision. Philosophers of science of different basic persuasions are still agreed that it is easier to write *merely* operational definitions than to work up operational definitions that lead to the possibility of powerful laws and theories in virtue of the fact that, while in a sense “conventional” or “arbitrary,” they still select out for mention the causally relevant aspects of the world—that they “slice the cake

rightly” or “carve Nature at its joints.” It seems, for example, unlikely that any reduction-sentences for such constructs as “expectancy,” “habit,” or “hunger” will in the long run lead to a predictively powerful theoretical structure if they are framed so as to ignore such nuisance variables as “fear,” “fatigue,” or “strength of competing exploratory response  $R_2$ .”

We have gone into this question in considerable detail, because this is the methodological orientation which leads us to prefer an implicit definition approach, via a whole system of postulates, to the more “direct operational reduction” attempted by Tolman, Ritchie and Kalish. It will be evident that the incomplete formalization which follows does *not* attempt to include guesses as to the causal laws defining such constructs as “exploratory drive” and “frustration.” Experiments aimed at testing the proffered system can, at best, claim to have (1) eliminated or (2) randomized, these unreduced factors. If any particular “test” experiment (pro or con) is then criticized as not having actually achieved the necessary elimination or randomization, such an *ad hoc* claim must then be itself formulated so as to permit an experimental test, new postulates being added for the purpose of defining the hypothesized nuisance factors in the given setting (and, if possible, more generally). We suggest that the present approach via implicit definition allows for these accretions with less violence to the basic theoretical structure than is the case with allegedly “direct, operational” definitions of the core constructs of the theory.

There seems to be no justification for leaping from Tolman’s near-zero degree of quantification to specific guesses at high levels of quantification. We have tried to carry him to around level (3) as a starter.

In considering an informal system such as Tolman’s, one must try to catch the flavor, get the kernel, isolate the *kind* of thing that seems to be stressed, and then say it in a reasonably specific confirmable form. In doing so, we must be prepared for the objection that “But that isn’t *quite* what Tolman means.” It is not easy to state precisely what

Tolman means, and in many cases he has not intended to express an opinion on points which an axiomatization must ultimately say something about. After all, it would be presumptuous to offer a formulation of somebody if he had already formulated himself. On the other hand, it would be surprising if an informal system of proposals, offered as a basis for a specified set of theorems, could be seen as in exact isomorphism with a more formalized set which actually did the job.

Anthropomorphic, intuitive, or common-sense content may be used freely at first, in arriving at formulations. This is especially true in working with Tolman's ideas since he himself proceeds thus in the context of discovery. Furthermore, the "naturalness" of a Tolman type of formulation for the layman, who usually finds it closer to his introspective notions than is the case with other theories, suggests that we might attend more closely to our common, naive view of behavior while trying to get a firmer grasp upon whatever theoretical truth lies in the expectancy approach. Such crutches are not, of course, part of the formulation arrived at, and are not to be considered relevant in evaluating the theory.

As the word *expect* is ordinarily used, it may be seen as involving three elements. When one says "I expect to find food in the refrigerator," he is referring implicitly to

- (1) What he sees, initially: "The refrigerator."
- (2) What he does when he sees it: "Open it."
- (3) What he sees after doing that: "Food."

In the vernacular, these three components are not equally stressed. Thus, in the example, the reference to behavior is suppressed or barely indicated by the verb "find." Verbalization of expectations concerning *loci* are especially likely to suppress the reference to behavior, confining themselves to a statement as to where something is. Verbalization of other types of expectancies are more likely to include a reference to behavior. Thus, "If I reach toward a (seen) pencil, I expect to feel it," "When you insult people they get angry at you,"

“You have to turn the knob before it will open,” “That rat expects food when he presses the lever.” It is possible that Tolman’s almost complete dependence upon the maze as an instrument of discovery has led him to carry over this easy suppression of the response term, and thus given him the difficulty in “getting to” behavior indicated by Guthrie’s well-known gibe that Tolman leaves the rat buried in thought. If this mistake (as we see it) is avoided, the getting-to-behavior need not present a problem any more acute for Tolman than for a non-expectancy theorist, as we shall try to show. The exclusive concern with the “learning a locus” kind of problem, where the stimuli are characteristic of *places* and the responses studied are *locomotions*, makes it easy to formulate an expectancy theory as “*S-S*,” a decision which generates endless difficulties in the ensuing development.

So when we characterize an organism’s expectancy, we have to indicate *what* he expects under *what* circumstances when he does *what*. The “basic theoretical element” in an expectancy theory differs from such an element as  $sH_R$  in that the expectancy involves somehow a *reference to what is expected*. In Hull, this expected thing is involved only “historically,” in the sense that it appears in the laws which state how  $sH_R$  grows (i.e., it is involved as the reinforcer of a specified  $sH_R$ ). But it is not involved in the characterization of  $sH_R$  itself, wherefore to identify the habit we need only two subscripts, not three. On the other hand, the basic theoretical element introduced below in our reformulation of Tolman does involve a reference to *R*. The problem of how to get to the response is solved by putting *R* in at the beginning, as in non-expectancy theories.

In what “language” are these three components to be characterized? Let us call them, the *elicitor*, the *response*, and the *expectandum* (from Latin gerundive “... to be expected”). Thus:

<i>Elicitor</i>	<i>Response</i>	<i>Expectandum</i>
$S_1$ : Choice-point stimulation	$R_1$ : Right turn	$S_2$ : Food-box

Obvious simplifications occur here, of course. “Turning right” is a reaction-class of variable members. The elicitor is a proximal stimulus

field bearing complex statistical relations to the physical situation. However, neither of these simplifications seems clearly more characteristic of Tolman's than of any other theory. A probabilizing treatment of "stimulus-elements" [such as Estes (26)] could be carried out within an expectancy frame as well.

The *expectandum* is also really a class of proximal stimulus fields, the members of which are dated occurrences. If one says loosely "an occurrence of *the* expectandum," he means the occurrence of a member of this class of proximal stimulus-fields which the organism realizes by entering into a certain physical relation to the named external object. Strictly, of course, this external relation also has a class-character. Often, as is also true of the elicitor, the expectandum will be alluded to in terms of physical objects. But this is always understood to be elliptical, and in any case where it makes a difference the term should mean a proximal stimulus class.

An *expectant* is a unique, dated occurrence of a sequence in which  $S_1$  (elicitor) occurs and arouses a central state ( $r_1s_2$ ). The aspect, "component" or phase of the central state designated by  $s_2$ , is an *expectate* (thus, also a unique, dated occurrence).

A disposition to have expectants of a certain sort (i.e., characterized by components  $S_1, R_1, S_2$ ) is an *expectancy*. It must be understood that these remarks are in the nature of preliminary explications, since they are manifestly not "operational definitions" in behaviorese. Such concepts as these are finally to be defined by the whole system of postulates in which they occur. Therefore, in the present volume they *remain* very incompletely defined.

$S_1$  (the elicitor) is obviously to be characterized in the usual stimulus language, and  $R_1$  in the response language. While  $S_2$  (expectandum) is a stimulus, in what language is its central representative ( $s_2$ , the expectate) to be characterized? Four "languages" are available:

- (1) Physical stimulus language.
- (2) Response language.
- (3) Physiological language.

## (4) Phenomenal language.

All but (1) are quickly eliminated as candidates. There is serious doubt as to whether there exists a (4) which is not really (1) with the subject reporting under special instructions; and in any case the rat hasn't told us his. Language (2) would miss the point by turning us back to a response theory (cf. role of  $r_G$  in non-expectancy formulations). Language (3) will be fine when we get it, although probably even then not the most useful for behavioral purposes. This leaves (1), which is what common-sense suggests anyway. "What do you expect around the comer?" calls for a description of the *objects* and *events* that will be there. It is also convenient to have the expectate in stimulus language since at some stage of the matter of "confirming expectancies" it will come up for quantification ("how *well* did the environment confirm the expectant?"), and it will be desirable if the same qualities and dimensions are available in discussing both the expectate and the confirming (or disappointing) environment.

Thus, to characterize the components of an expectancy we will ultimately require three sets of properties:

$S_1$ —Properties characterizing the elicitor.

$R_1$ —Properties characterizing the reaction-class (= response).

$S_2$ —Properties characterizing the expectandum.

The entire central state aroused by the elicitor  $S_1$  is the "expectant" ( $s_1r_1s_2$ ). While the denotatum of  $s_2$  is a neural event, *in its role as a hypothetical construct it is characterized by a set of stimulus-numbers*; just as the dimensions needed to identify the subscript  $R$  in  ${}_sH_R$  are *movement* dimensions, although  ${}_sH_R$  as a hypothetical construct itself is a central state or, as Brown and Farber would view it, a "calculational device" (4, p. 467). This sort of oddity seems to be the fate of a molar behaviorism and all one can do is try to get accustomed to it. There seems to be nothing methodologically wrong with it, since the presence and the quantification of a particular expectancy is inferred from the strength of a disposition to emit  $R_1$  in the presence of

$S_1$  provided the terminating stimulus  $S_2$  is currently valenced. That is, the central construct ( $S_1R_1S_2$ ) is tied to certain sets of data sentences about  $S_1$ ,  $R_1$ , and  $S_2$ . When these three symbols occur conjoined *in the brackets*, they (jointly!) denote an expectancy; which latter is a theoretical construct and hence *not* in data language. But when we want to know “*which* expectancy” is being talked about we need a notation which links us up to the ultimate coordinating definitions in data language. The expectandum, a proximal stimulus class, has stimulus dimensions; when the same capital letter occurs in the brackets, it does not—*could* not, for within the brackets it has no independent designating function at all.

The following skeletal set of “postulates,” with accompanying brief comments, illustrates how the expectancy-concept might be introduced. We have kept fairly close to Tolman’s language to avoid yet another new vocabulary. Each “law” contains sub-principles of varying independence. The term *monotonic* is to be understood throughout when functions are mentioned.

1. *Mnemonization*.—The occurrence of the sequence  $S_1 \rightarrow R_1 \rightarrow S_2$  (the adjacent members being in close temporal contiguity) results in an increment in the strength of an expectancy ( $S_1R_1S_2$ ). The strength increases as a decelerated function of the number of occurrences of the sequence. The growth rate is an increasing function of the absolute value of the valence of  $S_2$ . If the termination by  $S_2$  of the sequence  $S_1 \rightarrow R_1$  is random with respect to non-defining properties of  $S_1$ , the asymptote of strength is  $\leq$  the relative frequency  $P$  of  $S_2$  following  $S_1 \rightarrow R_1$  (i.e., a pure number). How far this asymptote is below  $P$  is a decelerated function of the delay between the inception of  $R_1$  and the occurrence of  $S_2$ .

*Comment:* This is one of the “basic acquisition postulates” for an expectancy theory, playing a role similar to Postulate 4 in Hull’s 1943 set. We have taken as a first guess the assumption that the valence of the expectandum affects *rate* of acquisition only, not asymptote. In common-sense terms, if an animal confirms an expectancy often

enough he will become “certain” of it whether it is valenced or not although, if it is not highly valenced, this process may take a very long time. The “absolute value” refers to the supposition that strong negative valences also yield rapid growth of expectancies, although presumably this will ultimately have to be greatly qualified by additional postulates concerning emotional “disrupting” effects.

Professor Tolman (personal communication) raises the question whether the third sentence, referring to the dependence of growth rate upon expectandum valence, is meant to imply that *zero* valence of  $S_2$  would mean zero growth. If so, he states that he would be in disagreement with it. We have left this question open deliberately in our phrasing of the postulate, to emphasize our own view that the necessity of reinforcement in the acquisition of “whatever is learned” is *not* the defining property of an *S-R* theory, nor is a denial of the necessity of reinforcement for learning the feature which defines an expectancy theory. We agree with Spence (97) that the issue “*S-R* versus expectancy theory” is logically distinct from the issue “reinforcement versus contiguity-as-sufficient,” and we are taking the first issue as crucial in what follows. Confining the discussion entirely to instrumental learning, it seems that four possibilities can be sketched *a priori* for the forms of learning theory, thus:

<b>Reinforcement issue</b>			
	<i>Reward necessary for the acquisition</i>	<i>Reward not necessary for the acquisition</i>	
<b>Issue as to form of basic learning construct (“what is learned”)</b>	<i>Habits: S-R connections</i>	Hull Miller Skinner (Type R)	Guthrie
	<i>Expectancies: SRS connections</i>	?	Tolman

If the parameters of the mnemonization function ( $SRS$ ) =  $f(n)$  are so

chosen that  $\frac{df}{dn} = 0$  when  $|V| = 0$ , we would have a theory to fill the lower left box. If  $\frac{df}{dn} > 0$  when  $|V| = 0$ , we have Tolman. The very fact that one can adopt the entire set of postulates and then make his guesses about this question *within* the expectancy frame seems to us the more reason for rejecting the "necessity of reward" as defining a major theoretical issue. The acquisition postulate might be, say, of the form  $(SRS) = M \left[ 1 - e^{-(a+bV)^n} \right]$  and it is hard to see the argument between "map" theories and "response" theories as somehow concealed in the question whether the constant  $a$  is exactly zero or instead is near zero. It is an interesting question why, historically, the above table has an empty cell.

The matter of asynchronism between  $S_1$  and  $R_1$  has been ignored except insofar as "close temporal contiguity" is required, since we are here dealing with instrumental (operant) learning only. The reference to the valence of  $S_2$  attempts to do some justice to the insistence of Leeper (53, p. 105) and Tolman (151, p. 150) that the animal's cognizing of an expectandum depends to some extent upon the latter's "importance" to the need-state, without prejudging the exact quantitative question just discussed. It will appear later that this is *not*, as a superficial analysis might suggest, tantamount to admitting the truth of some non-expectancy theory. Note also that the present form includes a reference to confirmation-frequencies  $< 1$ , since such cases are by far the commonest in "real life." Such less-than-invariable successions generate expectancies of lower strengths, which in turn lead (via the activation postulate below) to lower instantaneous strengths of response. Of course, if the probability of  $S_2$  is *non-random* with respect to some property of  $S_1$ , we are in effect setting up a discrimination between two elicitor subclasses  $S_1'$  and  $S_1''$ , and hence must reformulate the problem in terms of two expectancies. There must also be the possibility of a time-reference in the elicitor-term, to take care

of those temporal discriminations reflected in curves obtained under periodic reinforcement. Similarly, if there are non-defining properties of  $R$  with respect to which the  $S_2$  probability is not invariant, we deal then with a problem in response differentiation, i.e., these properties *become* defining properties for two new reaction-classes  $R_1'$  and  $R_1''$  and again we must begin speaking of two expectancies. In both cases the single postulate leads to a gradual drawing apart in strength. We have called the whole a principle of *mnemonization* following Tolman's distinction (1932) between this process and the other two cognition forms (perception and inference).

We have considered a radical addition to this postulate which would make the rate of expectancy-growth depend not merely on the valence of  $S_2$  but on the valences of all the approximately simultaneous stimuli; and such an addition may be found unavoidable. But after the *first* occurrence of a sequence  $S_1 \rightarrow R_1 \rightarrow (S_2 S^*)$  where the star superscript indicates that  $S^*$  is valenced, the expectandum  $S_2$  will have an induced cathexis (see 6: Secondary Cathexis below); and the response will, in addition, be receiving strength from the expectancy ( $S_1 R_1 S^*$ ). Hence experimental evidence for the suggested addition would be hard to get, requiring either a very sensitive operant or large numbers of animals.

2. *Extinction*.—The occurrence of a sequence  $S_1 \rightarrow R_1$ , if not terminated by  $S_2$ , produces a decrement in the expectancy if the objective  $S_2$ -probability has been 1.00, and the magnitude of this decrement is an increasing function of the valence of  $S_2$  and the current strength of ( $S_1 R_1 S_2$ ). Such a failure of  $S_2$  when  $P$  has been  $\doteq 1$  is a *disconfirmation* provided ( $S_1 R_1 S_2$ ) was non-zero. For cases where the  $S_2$ -probability has been  $< 1.00$ , if this objective probability  $P$  shifts to a lower  $P'$ , and remains stable there, the expectancy strength will approach some value  $\leq P'$  asymptotically.

*Comment:* This principle is in particularly poor shape, and we would perhaps do better to over-simplify by ignoring anything but expectancies near 1.00 at this stage. The principle is qualitatively

analogous to the usual extinction principle in non-expectancy theories. The notions of inhibition and energy-expenditure are not, however, introduced. To do so in this context would depart too much from the spirit of an expectancy theory. We are not primarily “tiring out” *responses*, we are instead “learning *that*  $S_2$  no longer follows unfaithfully.” The undoubted relevance of the work-parameters and the whole problem of spontaneous recovery must somehow be dealt with by an expectancy theory, but will better be fitted in somewhere else than here. Three different components of this question may be tentatively distinguished. First, we will have (as in this postulate) the reduction of the “cognitive” strength of the expectancy by a failure of the usual terminator. Secondly, we may assume that there are some work-related consequences of movement similar to Hull’s  $I_R$ , which depress effective reaction-potential without altering expectancies (just as it does not alter  $sH_R$  in Hull). Thirdly, there is a possibility of an effort-dependent *parameter* being required in the equation for  $sE_R$ , quite apart from any question of “accumulating consequences” of work. Such a parameter would not be confined to extinction effects but would also enter into the ordinary determination of a rate of responding, whether the schedule of reinforcement were periodic, aperiodic, or continuous (cf. 91, p. 202).

If we had enough leads on “anxiety” to anchor it more adequately on the dependent variable side, we would be tempted to insert here some kind of “Law of Disappointment: The occurrence of a disconfirmation of ( $S_1R_1S_2$ ) if  $S_2$  has positive valence arouses anxiety, the amount of anxiety being an increasing function of the valence of the expectandum  $S_2$ .”

When the initial  $P < 1$ , the decline from  $P$  to  $P'$  presents special problems. In the case of a schedule of periodic reinforcement we have the option of (a) speaking of a family of expectancies, the elicitor of each being made to include time, rate, or serial position as a “stimulus” variable (cf. 90, pp. 263-265); or (b) speaking of some sort of generalization gradient from the strongest value of a single

expectancy. But on an aperiodic schedule these possibilities are not available, at least in any obvious way. Common sense would say, "If the rat gets a pellet every time, a disconfirmation will be noticed. Similarly, although with a slight vagueness because he cannot differentiate time precisely, the rat notices a failure to deliver at the end of four minutes. But if the administration of  $S_2$  has been random, he must be kept on the new schedule for a time before he lowers his anticipations." These problems remain to be taken up constructively, but the skimpy state of present evidence hardly justifies an attempt as yet.

Another interesting possibility is that of inducing negative cathexes via the disconfirmation operation. There might be some principle to the effect that "Disconfirmation of an expectancy ( $S_1R_1S_2$ ) by terminating the sequence  $S_1 \rightarrow R_1$  by a non-valenced stimulus (or, less-valenced)  $S_k$  instead of  $S_2$  induces a negative cathexis in  $S_k$  provided that the expectandum  $S_2$  had a positive valence at the time of the disconfirmation. The increment given to this negative cathexis (or, more generally, the decrement given to this cathexis) is an increasing function of the valence of  $S_2$ , and the negative *valence* given  $S_k$  by this negative cathexis will covary with subsequent change in the need which gave  $S_2$  its valence." Such a principle might be used to deal with the Kendler-Mencher data (49), where each sight of food in the "wrong" cup meant disappointment of the water-expectancy. One thinks also of the oft-cited "disappointment" of Tinklepaugh's (109) monkeys, or Crespi's work on shifting to smaller maze rewards (15, 16). As stated in the postulate, the idea is that the *single* disconfirmation produces a decline in a near-unity expectancy. It also produces a decline in a weaker expectancy which is, however, approaching unity as an asymptote under a continuous schedule ( $P_2 \doteq 1.00$ ). But in this second case the absolute decrement is less. In both cases the decrement also depends on the valence of  $S_2$ . If we have a  $P_2 < 1.00$  to start with, the single "disconfirmation" cannot play such a role, as we have just discussed.

3. *Primary stimulus generalization.*—When an expectancy ( $S_1R_1S_2$ )

is raised to some strength, expectancies sharing the  $R$  and  $S_2$  terms and resembling it on the elicitor side will receive some strength, this generalization strength being a function of the similarity of their elicitors to  $S_1$ . The same is true of extinction of  $(S_1R_1S_2)$ .

*Comment:* This again plays a role similar to that which generalization plays in a non-expectancy system. It seems repugnant to common sense in one way, namely, that we do not seem to expect  $S_2$  when  $S_1'$  is "similar," in certain cases when we can "tell it apart from"  $S_1$ . But this common-sense objection seems to be mainly true of our expectations regarding *places*. Perhaps some additional assumptions as to long-term breaking down of generalization tendencies, i.e., a reduction in the parameters themselves, may be needed.

The use of the non-committal term *similarity* instead of a reference to units on a particular stimulus continuum is deliberate. The postulate is intended to cover a host of specific perceptual laws, each to be separately investigated. No reference to the acceleration of the generalization function is here appropriate; the many sorts of abscissa-variables which will occur in these gradients of "similarity," and a certain arbitrariness in their quantification, make untimely a reference to the sign of the second derivative of such gradients.

Tolman has been accused of referring vaguely to "laws of perception" when the problems of generalization and stimulus equivalence come up. To be sure, this is what he does; but it is not clear just who is in a position to criticize him for it. There is a serious question as to whether anyone else does more than this (validly) for any but the simple, one-dimensional case. For this simple case, a decay function against abscissa stimulus values (whether in absolute units or j.n.d.'s) may be substituted in the above postulate by the reader who wishes to do so.

4. *Inference.*—The occurrence of a temporal contiguity  $S_2S^*$  when  $(S_1R_1S_2)$  has non-zero strength, produces an increment in the strength of a new expectancy  $(S_1R_1S^*)$ . The induced strength increases as a decelerated function of the number of such contiguities. The asymp-

tote is the strength of  $(S_1 R_1 S_2)$  and the growth rate is an increasing decelerated function of the absolute valence of  $S^*$ . The presentation of  $S_2$  without  $S^*$  weakens such an induced expectancy  $S_1 R_1 S^*$ . The decrement is greater if the failure of  $S^*$  occurs at the termination of the sequence  $S_1 \rightarrow R_1 \rightarrow S_2$  than if it occurs as a result of presentation of  $S_2$  without  $S^*$  but not following an occurrence of the sequence.

*Comment:* We believe that this is the sort of postulate whose presence contributes heavily to the identification of an expectancy theory as such. The term "expectancy," while it occurs in the Law of Mnemonization, cannot immediately be distinguished in its empirical consequences by its role there, from the alternative "habit" or " $S$ - $R$  bond." From reading the law of mnemonization one could see no good reason for including  $S_2$  as part of the "expectancy." For if the valence of  $S_2$  depends jointly on *need* and *cathexis*, and the latter is not being experimentally manipulated, then the only really relevant use of the valence of  $S_2$  is in the acquisition postulate, where we can often make  $S_2$  play the role of the "reinforcer" and drop any reference to it in characterizing *that which* is being thus strengthened.

In such a case we would have reduced  $(S_1 R_1 S_2)$  to  ${}_S H_R$ . But the role an expectancy takes on by virtue of the occurrence of  $S_2$  in the Inference Postulate, *and by the kind of strengthening operation this makes possible*, permits it to behave in the entire system in a new way. This new way makes the "basic bond"  $(S_1 R_1 S_2)$  correspond in its properties more to what Tolman (and the layman) seem to mean by an "expectation," and less to the usual psychological definition of a "habit."

The common sense of the situation is obvious. If one has learned to expect that he can bring about a certain situation by doing so-and-so; and then (subsequently) he finds that a new element is to be found in that situation; he "infers" that he will be able to bring about the presence of this new element by doing so-and-so. It may be objected that this differs from the idea of secondary reinforcement of an  ${}_S H_R$  only in the time-order of the two acquisition procedures. In the usual case covered by the "secondary reinforcement" concept, we first pair a

known reinforcer  $S^*$  with a neutral  $S_2$ ; this pairing is said, by the principle of secondary reinforcement (a postulate?), to confer the "reinforcing property" upon  $S_2$ . When we now terminate the sequence  $S_1 \rightarrow R_1$  by presenting  $S_2$ , the basic principle of reinforcement can then be invoked to derive the consequent strengthening of  ${}_sH_{R_1}$ . Expectancy theory proceeds almost identically by use of the principle of secondary cathexis. In the other time order, we first run off the sequence  $S_1 \rightarrow R_1 \rightarrow S_2$  repeatedly (observing no increase in the tendency of  $S_1$  to elicit  $R_1$ ). Following this, we pair  $S_2$  with the valenced  $S^*$ , but not following runnings-off of the sequence. This pairing confers the "reinforcing property" upon  $S_2$ ; but there is no principle which we can now apply retroactively to strengthen the bond between  $S_1$  and  $R_1$ . "Only" the time-order is what makes the big difference. *One* order is easy for both expectancy and nonexpectancy formulations, given a postulate or corollary of secondary reinforcement. The *other* order is practically impossible for non-expectancy theories except by an introduction of additional responses as mediators. After all, there is a sense in which time-relations are the *essence* of all learning theories, since they all make use of the basic relations of contiguity and temporal succession. For instance, if an association of topographically unrelated elements took place on the basis of a "contiguity" involving a five-year interval, all theorists would be equally stupefied. The time-relations are the crux of learning situations, so an objection that the difference is "merely one of time-order" cannot be sustained.

We have ignored here the generalization problem. It must be supposed that contiguities of  $S^*$  and some fourth stimulus  $S_2'$  will induce strength in  $S_1R_1S^*$  provided that  $S_2$  and  $S_2'$  are sufficiently similar. Perhaps a Law of Primary Generalization should be stated for expectanda as well, as follows:

5. *Generalized inference*.—The occurrence of a temporal contiguity  $S_2S^*$  produces an increment in the strength of an expectancy  $S_2R_1S^*$  provided that an expectancy  $S_1R_1S_2'$  was at some strength and the expectandum  $S_2'$  is similar to  $S_2$ . The induced strength increases as a

decelerated function of the number of such contiguities. The asymptote is a function of the strength of  $S_1 R_1 S_2'$  and the difference between  $S_2$  and  $S_2'$ . The growth rate to this asymptote is an increasing decelerated function of the absolute valence of  $S^*$ .

6. *Secondary cathexis*.—The contiguity of  $S_2$  and  $S^*$  when  $S^*$  has valence  $|V|$  produces an increment in the absolute cathexis of  $S_2$ . The derived cathexis is an increasing decelerated function of the number of contiguities and the asymptote is an increasing decelerated function of  $|V|$  during the contiguities, and has the same sign as the  $V$  of  $S^*$ . The presentation of  $S_2$  without  $S^*$ , or with  $S^*$  having had its absolute valence decreased, will produce a decrement in the induced cathexis of  $S_2$ .

*Comments*: This is “secondary reinforcement,” and in Tolman’s 1950 language some mixture of the learning of cathexes and equivalence-beliefs. Since we have been unable to distinguish these two to our satisfaction, we have put them into one principle. They also appear somehow fused in the principle of Elicitor-Cathexis below.

The induced cathexis presumably cannot be said to approach an asymptote which depends only upon the *cathexis* of the primary stimulus  $S^*$ , since under low drive we would expect that contiguity to confer less induced value upon  $S_2$ . Hence, we have made this asymptote hinge upon the *valence* (cathexis-need combination) of  $S^*$ . But the reasoning here is not so evident as is sometimes thought. A more daring guess would be to rephrase the asymptote reference and say “... and the asymptote is that of the inducing cathexis,” (*not* valence). This would mean that if a neutral stimulus is repeatedly paired with a strongly cathected goal-object but under low need, a subsequent rise in need should give the (now) secondarily cathected stimulus the same large valence it would have had if the pairing had taken place under high need.

7. *Induced elicitor-cathexis*.—The acquisition of valence by an expectandum  $S_2$  belonging to an existing expectancy ( $S_1 R_1 S_2$ ) induces a cathexis in the elicitor  $S_1$ , the strength of the induced cathexis being

a decelerated increasing function of the strength of the expectancy and the absolute valence of  $S_2$ .

8. *Confirmed elicitor-cathexis*.—The confirmation of an expectancy ( $S_1R_1S_2$ ), i.e., the occurrence of the sequence  $S_1 \rightarrow R_1 \rightarrow S_2$  when ( $S_1R_1S_2$ ) is of non-zero strength, when  $S_2$  has a positive valence, produces an increment in the cathexis of the elicitor  $S_1$ .

This increment in the elicitor-cathexis by *confirmation* is greater than the increment which would be *induced* by producing a valence in  $S_2$  when the expectancy is at the same strength as that reached by the present confirmation.

*Comment:* Some such postulate seems in order considering the reinforcing properties acquired by discriminative stimuli (cf. 90, pp. 245-253; 43, p. 236). The distinction between induced and confirmed means that, e.g., choice-point stimuli acquire more cathexis when hungry rats run the T-maze to food, than such stimuli would acquire if rats are made hungry after a large number of satiated runs, even if enough of the latter have occurred to bring the expectancies themselves to a level comparable to those acquired in the first case.

9. *Valence*.—The valence of a stimulus  $S^*$  is a multiplicative function of the correlated *need*  $D$  and the *cathexis*  $C^*$  attached to  $S^*$ . (Applies only to cases of positive cathexis.)

*Comment:* Such a principle is empirically useless without the whole mass of imbedding material required to elucidate the relation of cathexes and needs. The need-concept may be introduced solely by reference to the two basic facts of (a) strengthening responses by manipulation of their stimulus-consequences, and (b) experimental relevance of maintenance schedule with respect to the stimuli used. If such a “non-physiological” approach is followed there are, strictly speaking, as many “needs” as there are goal-object-classes found to be not completely interchangeable and which can be shown to be related to a maintenance-schedule (cf. 63). If this procedure is followed, we have two other principles to go on, once a class of stimulus-situations has been identified as a goal-situation and the strength of the assoc-

iated need defined by its maintenance schedule.

10. *Need strength*.—The need ( $D$ ) for a cathected situation is an increasing function of the time-interval since satiation for it.

Upon present evidence, even basic questions of monotonousness and acceleration are unsettled for the alimentary drives of the rat, let alone other drives and other species. There is no very cogent evidence that all or most needs rise as a function of time since satiation, although this seems frequently assumed. The notion of satiation itself, even in connection with “simple” alimentary drives, presents great difficulties.

11. *Cathexis*.—The cathexis of a stimulus situation  $S^*$  is an increasing decelerated function of the number of contiguities between it and the occurrences of the consummatory response. The asymptote is an increasing function of the need strength present during these contiguities. (There may, however, be some innately determined cathexes.)

*Comment:* Such a principle would seem to be required unless we assume more biologically given cathexes for exteroceptive stimuli than seems plausible. If correct, it makes necessary an ample opportunity to acquire the cathexes in pre-experimental exposures, before an experiment on latent learning is begun (cf. 13, 14).

Perhaps any special effort to introduce postulates concerning goal-objects should be avoided, and the whole problem handled by implicit definition in which the consummatory response is itself merely treated as the response-term of an expectancy. This final expectancy has, e.g., food as the elicitor, eating as the response, and immediate gustatory and proprioceptive consequences as expectanda. These latter are highly cathected, the confirmation is immediate and invariable, and the number of confirmations is tremendous. The primary cathexis is then assumed to lie in the feel, taste, and smell of the food in the mouth, and the proprioceptive consequences of chewing and swallowing. By the principle of Elicitor-Cathexis, the sight, smell, or touch of food not yet ingested ought to acquire a strength of cathexis very close to the primary. Here, of course, one must make a guess as to how far

back is “primary,” and whether what is primary in the Law of Effect is stimulation, cessation of stimulation, the necessary behavior-supports to *respond* in a certain way, the alteration of a central state via the bloodstream, or some composite of these. Tolman has referred to certain “to-be-got-at physiological quiescences,” with respect to which all other goal properties are ultimately docile. In the absence of any clear evidence, we have formulated the expectancy principles in terms of a stimulus-reinforcement view. That is, food has been treated as rewarding by being seen, smelled, or touched. We spoke of  $S^*$  and  $S_2$  being contiguous, and made no mention of a consummatory response. Beliefs may differ as to the reinforcing value of the consummatory response itself. However, whether the *act* of chewing (as an efferent event) or the immediate sensory consequences of the act, is what gives it its reward-value, makes little difference for the present task. Since the two are (barring surgical tricks) necessary and sufficient conditions for each other, we can refer merely to the “consummatory response” and get to all anterior (and presumably derived) cathexes from there.

12. *Activation.*—The reaction-potential  ${}_sE_R$  of a response  $R_1$  in the presence of  $S_1$  is a multiplicative function of the strength of the expectancy ( $S_1R_1S_2$ ) and the valence (retaining sign) of the expectandum. There are momentary oscillations of reaction-potential about this value  ${}_sE_R$ , the frequency distribution being at least unimodal in form. The oscillations of two different  ${}_sE_R$ 's are treated as independent, and the response which is momentarily “ahead” is assumed to be emitted.

*Comment:* From here on in the intervening variable chain there need be little or no difference between expectancy and non-expectancy theories. As in the case of non-expectancy theories (cf. Postulate VIII, Corollary v in 40, p. 8), a problem arises in connection with multiple expectancies having the same response term. We have made no attempt to consider the formalization of the overlapping case. One sense in which the “field” emphasis might seem intimately connected with the expectancy view (contrary to our separation of these issues)

involves this question. For one might set up (conceptually) as many separate expectancies as there are stimuli in the goal-box. That is, there is an expectancy of wood, of white color, of sawdust, of a food-cup, and so on. A disturbing arbitrariness arises here, since by “cutting it fine” we could find ourselves considering indefinitely large numbers of expectancies, each coordinated to some describable aspect of the physical situation terminating  $S_1 \rightarrow R_1$ . It might be inferred from this that some substituting of the “field” as expectandum will avoid the difficulty, and hence that the “field approach” is more intimately related to an expectancy theory than to  $S$ - $R$  formulations. However, the same appearance of arbitrariness occurs on the elicitor side, and hence for non-expectancy theories. Why are there not as many *habits* being conditioned in the T-maze as there are physically describable “stimuli” at the choice point? And, since the rat does not always sample all of the available (external) stimulus energies on every trial, these habits must be presumed to be growing at different rates; presumably they also have different asymptotes because of the different modalities and stimulus intensity dynamisms involved. We do not mean to underestimate the difficulties merely by pointing to their occurrence elsewhere, of course. But the present interest is not in those difficult problems which are shared by expectancy and non-expectancy theories. What is actually done in most contemporary experimental speaking about  $S_1$  is to characterize the apparatus and let it go at that. Skinner’s emphasis on the generic nature of stimulus and response stems from his awareness of the problem. Interest is centered on actual stimulus components of the starting-box or choice-point only when they are playing a discriminative role—in which case we would talk similarly about *two* expectancies (with different elicitors) in the present formulation also.

A law of threshold of  $sE_R$  to produce response, laws relating response measures (amplitude, latency, response-probability), and a more detailed law of oscillation might be similar to or identical with Hullian principles.

Some principle of chaining may be in order, beyond what is yielded by, e.g., the elicitor-cathexis principle. But the statement of a chained-expectancy postulate is pointless until some activation-postulate for chains is also stated. We might, for example, choose to concern ourselves only with how to calculate the strength of the initial member since, as soon as it occurs, the problem is re-presented. (We ignore the "prediction" of any  $sE_R$  until we are in the presence of *its* elicitor.) But the problem is to formulate a law regarding the manner of summation of the valences attached to expectanda later in the chain. Looking at Tolman's fused balloons on page 147 of the *Purposive Behavior in Animals and Men*, one feels the need for some postulate which will attach the subsequent expectanda of a chain to the response term of each member expectancy. Presumably this strength depends on the strengths of the links (perhaps not stronger than the weakest), and is greater when the chain has been formed by the running off of the chain *sequence* than when it is only induced by the separate confirmation of its elements. But obviously all of this is the sheerest speculation and we have not even sketched a postulate.

For a few sample derivations from these principles for the T-maze case, the reader is referred to (59). Here let us return briefly to the possibility of deriving "multiple trackness" for the lever-pressing case considered above in connection with the response problem. The expectancy ( $S_{\bar{h}}R_lS_{\bar{h}}$ ) where the bar over  $h$  or  $l$  means home-cage or experimental lever *down*, has been strengthened by the pre-experimental history in the home-cage (Postulate 1, Mnemonization). Since the home-cage bar for operating food-dispenser and the lever in the experimental box are similar, the expectancy ( $S_lR_lS_{\bar{h}}$ ) is strong by Postulate 3, Primary Generalization. In the box, repeated pairings of "lever down" ( $S_{\bar{l}}$ ) with food-pellet presentation ( $S_{j^*}$ ) generate an expectancy ( $S_{\bar{l}}R_lS_{j^*}$ ), by Postulate 5 (Generalized Inference), since  $S_{\bar{l}}$  resembles  $S_{\bar{h}}$ . Any topographic difference between teeth-pressing in the home-cage and that in the box is assumed to be well within the inductive range of the class  $R$ . If this were not assumed, the substi-

tution of  $(S_l R_l' S_{\bar{l}})$  for  $(S_l R_l S_{\bar{l}})$  would, of course, require an additional postulate of response induction. The important point here is that the derivation is mediated by the inference postulate, and hence involves a learning construct which notationally refers not only to the elicitor and the response but to a third element, the expectandum, as well. It is via the reference to  $S_{\bar{l}}$ , "lever down," that the derivation can be carried through. The formal structure here mirrors our informal notions of an expectation, and would be paraphrased roughly thus: "In the home-cage, the rat has learned that when he sees a lever-like object, he can get it down with his teeth; in the box he learns that when the lever is down he gets food. He infers that he can get food by using his teeth." An alternative derivation would invoke Postulate 6, Secondary Cathexis, proceeding via the acquisition by  $S_{\bar{l}}$  of a cathexis. These two are not, of course, opposed in their outcome or in any way incompatible. It is perhaps worth noting that, as is often the case, not *all* of the "common-sense" notion of the situation appears in the more formal deductions. The derivation actually makes no use of the cage-expectancy  $(S_h R_p S_{\bar{h}})$ , and its resemblance to  $(S_l R_p S_{\bar{h}})$ . The occurrence of the sequence  $S_l \rightarrow R_p \rightarrow S_{\bar{l}}$  serves only to bring about the juxtapositions  $S_{\bar{l}} S_f^*$  needed for the generalized inference postulate. The anthromorphic linkage between "there are two ways to get a lever down" and "the first way gets food" to "then so also will the second way" finds no place in the formal reconstruction. Whether this would turn out to be the case for all instances of multiple trackness, or even whether all experimental instances of the latter could be derived from the present postulates, we shall not consider here.

## VI. APOLOGY TO TOLMANITES AND OTHERS

We are aware that certain objections can be made to the foregoing as even a partial formalization of Tolman's proposals. As was pointed out in the preceding, the question "Does such-and-such a formulation say what Tolman had in mind?" is, strictly speaking, unanswerable,

since answering it involves the comparison of a semi-definite set of proposals with an even less definite set. Perhaps an informal consideration of the spirit of Tolman's approach is the most we can offer beyond what has been said. Obviously what follows can make no claims to rigor, but we hope that certain intuitively based or common-sense resistances to such a formalization will be somewhat reduced.

The most striking departure from Tolman's general line is probably our stress on the *response*. The basic theoretical element, the whatever-it-is that receives an increment from the occurrence of the sequence  $S_1 \rightarrow R_1 \rightarrow S_2$ , while called an *expectancy*, nevertheless has a response reference in it from the first. We are prepared to admit that this is a radical departure from the current form of Tolman's thought, although not so much from that of the 1932 book. (Cf. 128, pp. 10-12, 82.) However, in other places, e.g., p. 136, the explicit reference to *behavior* has almost vanished except for the phrase "commerce with." It is difficult to judge the book on this point because of Tolman's constant use of "means-end-relation," a phrase which does not clearly exclude or include a response-reference. The Glossary does not help (p. 451), since the word "direction" appears in the definition and the definition of this term (p. 441) is similarly interpretable in both ways. On the whole, the *map* emphasis and the failure to include a specific response-reference is already detectable in 1932. But we are not persuaded that this aspect is as definitive of Tolman as, say, the kind of acquisition assumption involved in our Inference Postulate, or the kind of energizing assumption involved in our Activation Postulate.

Now one of the basic difficulties with Tolman's system, admitted at times by Tolman himself and stressed by his critics, is the difficulty of getting from "knowledge" to "action." His tentative concession to Guthrie as to how a motor pattern gets acquired (151, pp. 153-154) is not, insofar as we understand it, fitted into his general cognitive framework with any great care. Taken literally, it almost seems incompatible with the rest. Tolman seems to be agreeing with Guthrie that "Any response (i.e., any movement) which goes off will ... get

conditioned on a single trial to whatever stimuli were then present” (p. 153). At face value such a remark concedes the case to the *S-R*-contiguity theorist. Although this cannot, of course, be Tolman’s intention, he can hardly be said to have resolved the conflict by such a general remark as “... such a learning of motor patterns is of necessity always imbedded in a larger goal-directed activity ...” (p. 154). “Imbedded in” is not a precise methodological expression and as here used bridges some very serious axiomatic gaps.

Roughly speaking, there are two approaches to the response-cognition problem. One is to formulate the hypothetical cognition without reference to a response term or dimension, so that the characterization of an animal’s cognition takes the form of some sort of quasi-propositional event occurring within the animal. The elements of such an inner event are alluded to by quasi-data words, i.e., words which (out of the psychological context) would denote environmental objects and their relations, as discussed above. This approach is indicated by the use of such metaphors as “map,” and the increasing use of this metaphor by Tolmanites we interpret to mean that they favor such an “environment-referential” approach. The phrase “to learn the location of” stresses the same thing, in contrast to the phrase “to learn what response leads to.” Now it appears to us that what we may call the “environmental” kind of expectancy theory would present great difficulties in behavior linkage *even if the axiomatization of the cognition-acquisition process were well-developed*. Presumably a rat must have some additional expectancies regarding “what locomotions take me to  $S_2$ ” or “what manipulations by me in the presence of  $S_1$  will produce  $S_2$ .” Otherwise, his purely *map*-expectancies regarding the “objective” relations of  $S_2$  to other  $S$ ’s would remain, so to say, behaviorally irrelevant. Tolman has not addressed himself seriously to this question, but has left it at a rather common-sense level. If the rat, as a reasonable being, knew what led to what, he would do so-and-so. But if we try to anticipate the difficulties, it seems likely that an expectancy formula which *does* contain the response as an integral part of

the strengthened element will have ultimately to be inserted as a sort of subtheory in order to get from an “environmental” expectancy to the behavior. If we are right in this, the obvious suggestion would be to start afresh and put  $R$  in somehow at the beginning. If this means “watering down the difference between Tolman and Hull,” so much the better for it. Of course, the complicated problems of response-equivalence and the specification of the reaction-class  $R_1$  are still with us, shared with the competitor theories.

This “response” form of expectancy theory may retain a good deal of the (non-metaphorical) properties suggested by the map-metaphor. But the map-properties will have to emerge as derivative properties, the basic elements still involving response reference. This remark must, of course, *not* be taken to mean that responses in the sense of effector-activity-classes are the “building-blocks,” links, or units out of which expectancies are *physically* constituted—that the elicitation of an expectancy always involves the occurrence of an effector response, however minimal. That would be  $S$ - $R$  theory again, of course. We mean merely that the characterization of a complex cognitive map will involve a reference to its constituent expectancies, and these expectancies are themselves characterized notationally by response *reference*. Such a statement is quite different from any statement that reduces cognitions to implicit movements (cf. 59, p. 56.)

The experimental use of even relatively simple space-and-locomotion set-ups may already involve too great complications for the present primitive state of expectancy theory. In the absence of carefully controlled studies of the effect of early, non-experimental space-traversing experience in the rat, we must not neglect the possibility that appropriate behavior in direction- and locus-learning set-ups is itself a high-level, derivative phenomenon, actually unsuitable for the elucidation of the primary laws of expectancy in spite of its apparent simplicity (to human beings with very similar spatial-locomotor histories!). In order to give some concreteness to our suggestion that the map-properties might be viewed as high-level, derivative con-

sequences of primary response-referring expectancies, the following skeleton analysis of a "simple" spatial inference is offered. Needless to say we do not offer the "postulate" as even a guess at the truth, but only as illustrative. In what follows, reference to an expectancy merely designates its terms, so that it may in fact be of zero strength. If an expectancy has been strengthened by the occurrence of the sequence, it is confirmed, otherwise induced. The class relationships are thus:

$$\text{Expectancies} \left\{ \begin{array}{l} \text{Zero} \\ \text{Non-zero} \left\{ \begin{array}{l} ? \text{ "Primary" or "native" } \\ \text{Confirmed (e.g., by mnemonization)} \\ \text{Induced (e.g., by inference)} \end{array} \right. \end{array} \right.$$

Before we can even state the single postulate, we need a rather cumbersome set of definitions:

1. A set of  $k$  expectancies, ( $k - 1$ ) of which belong to a chain and the  $k$ th one has elicitor and expectandum terms stimulus-equivalent with the elicitor and expectandum of the chain, is a *circular set*. The chain and its alternate expectancy are *equivalent*.
2. If both the chain and the equivalent of the circular set have been confirmed, we have a *confirmed circular set*.
3. Two circular sets in which
  - (a) The number of expectancies is the same
  - (b) The response-terms in corresponding positions in the two sets are so similar in topography that near-perfect primary induction would occur between them (or: that the amount of primary induction between a member of  $R_j$  belonging to the first circular set *and* a member of  $R'_j$  homologous to it in the other set, is as great as that between two randomly chosen members of  $R_j$ , on the average)
 are *isomorphic circular sets*.
4. Two or more isomorphic circular sets which are confirmed belong to a *confirmed subfamily* of circular sets.

5. The class of all isomorphic circular sets some of which are a confirmed subfamily is a *confirmation-family*.
6. The *diversity* of a confirmation-family is some increasing function of the differences between stimuli which are in homologous positions in the circular sets of the confirmed subfamily. (The term “difference” here is not question-begging, since the differences of these stimuli will have been investigated directly, outside of this context, through studies of primary stimulus generalization and equivalence.)

With all this vocabulary we can now state a “postulate”:

Postulate 1001: The strengthening of a chained expectancy which belongs to a circular set will induce strength in the equivalent expectancy of the set, provided that the circular set belongs to a confirmation-family. The amount of induced strength is an increasing function of the chain strength, of the strengths of the expectancies in the confirmed subfamily, and of the number and diversity of the confirmed subfamily. It is a decreasing function of the variability of strengths over the confirmed family.

Let us illustrate this ponderous business by something concrete. A rat has had the following set of expectancies very consistently confirmed in his home cage:

I.	{	$H_1$	$S_1$ : distant metal tag. $R_1$ : forward.	
			$S_2$ : close tag, cage wall.	Confirmed
	{	$H_2$	$S_2$ : close tag, cage wall. $R_2$ : left turn.	
			$S_3$ : water nozzle.	Confirmed
	{	$H_3$	$S_1$ : distant metal tag. $R_3$ : 45° left turn.	
			$S_3$ : water nozzle.	Confirmed

$H_1, H_2, H_3$  are a confirmed circular set.  $H_1H_2$  is a chain, and  $H_3$  is its equivalent.

Now suppose during a test breaking phase this rat has the following circular set confirmed (equally consistently, although with a lower total frequency) while exploring a table-top:

II.	{	$H_{13}$	$S_{13}$ : north wall. $R_1'$ : forward.	
		$S_{14}$ : edge of table.	Confirmed	
	{	$H_{14}$	$S_{14}$ : edge of table. $R_2'$ : left turn.	
		$S_{15}$ : cup at NW corner.	Confirmed	
	{	$H_{15}$	$S_{13}$ : north wall. $R_3'$ : 45° left turn.	
		$S_{15}$ : cup at NW corner.	Confirmed	

$H_{13}$ ,  $H_{14}$ ,  $H_{15}$  are a confirmed circular set.  $H_{13}H_{14}$  is a chain and  $H_{15}$  is its equivalent.

I and II are each confirmed circular sets. Since the responses in corresponding positions are similar ( $R_1$  to  $R_1'$ ,  $R_2$  to  $R_2'$ , and  $R_3$  to  $R_3'$ ) they are isomorphic circular sets. Since each is confirmed, they define a confirmed subfamily. Actually, of course, not two but a considerable number of such sets are confirmed by any normal organism in the course of its non-experimental history, even in a rather restricted environment.

Now we put a rat in a "new" situation, teaching him two expectancies:

$H_{49}$	$S_{49}$ : wood floor. $R_1''$ : forward.	
	$S_{50}$ : iron cross-piece.	Confirmed
$H_{50}$	$S_{50}$ : iron cross-piece. $R_2''$ : left turn.	
	$S_{51}$ : goal-box.	Confirmed

Now consider an expectancy

$H_{51}$	$S_{49}$ : wood floor. $R_3''$ : 45° left turn.	
	$S_{51}$ : goal-box.	Unconfirmed

$H_{49}$ ,  $H_{50}$ ,  $H_{51}$  are a circular set, and they belong to a confirmation family. Hence, by the postulate,  $H_{51}$  will have induced strength. Since the family is well-confirmed and the diversity is great (the homologous elicitors and expectanda being very different), the induced strength should be high. The rat should "expect" goal-box 45° to the left, even though he has never made such a response to get it. The important thing to see is that the "inference" is mediated by a some-

what complex relational fact involving both response-resemblance and a confirmation-history, rather than any straightforward, simple stimulus generalization or response induction.

One may be struck by the cumbersomeness of this schematizing, but we have been unable to reduce it. Actually the formulation is grossly over-simplified and elliptical at several points, e.g., in a reference to the "strength of the chain," where variation among the several link strengths would presumably make a difference. This "simple" kind of appropriate behavior will almost certainly require a *more* rather than a *less* complicated mediation than the present one. If the reader is inclined to react (as we were) by saying, "All that, to say that if a rat learns to take an oblique angle as a short-cut, he'll try it in new situations," that is *precisely* the point of our example. Anyone attempting a development of this "obvious good sense" shown by the rat, reducing the number of definitions needed to frame a postulate and yet aiming at some reasonable approximation to rigor, will be convinced, we think, that we are not here straining at a gnat. And it seems to us that, *if* such a postulate were to be made the basis for a rat's "inference" and his resulting  $sE_R$  at  $45^\circ$ , the scientifically important components of the map-metaphor would be contained therein. It should be unnecessary to add that such a postulate would hardly be confined to the "map" case. The term *response* occurs in the postulate without restriction, and presumably such a high-order sort of generalization effect, if postulated, would apply to *social* and *manipulative* as well as *locomotor* behavior. One thinks of Harlow's "learning sets," and of Lashley's famous monkey shifting to the other hand, or his rat running over the maze top "toward the goal."

This development may be expected to arouse criticism from non-expectancy theorists, and one which it is easy to make against the whole formulation suggested in the foregoing pages. One gets the feeling that too much of what we "know" at the ordinary, common-sense level is being put into the basic principles. Since our own bias is mainly on the nonexpectancy side, we have had that feeling almost

constantly. However, Tolman's system *is* somehow closer to lay thought, and this in itself is surely no basis of rejection. Everyone would be interested in knowing more about intermediate steps, but what if the postulates were "correct" and the intermediate steps were in the central nervous system? When one feels the urge to request the "mediating processes" underlying an expectancy-formation or (especially) the inference postulate, he should ask himself: Do I require this because my *S-R* orientation gives me a conviction that any such gross-behavior law really involves a chain of effector-events mediating it? If the answer to this question is affirmative, it means that the postulate is not being taken as it is offered in an expectancy theory. If the answer is negative, the critic must presumably be demanding a reduction to the neurophysiology.

It is in this sense that the Tolman-Hull controversy involves a centralism-peripheralism issue (cf. 66). In one rather trivial use of the words, any behavioral system is both a "centralism" and a "peripheralism." The confirmation-basis is sentences about an organism's movements, and movements are the data which the science is *about*. Hence, Tolman is a peripheralist, because he is a behaviorist. On the other hand, hypothetical constructs utilized to mediate the stimulus-response laws are universally assumed to have their physical locus in the animal's brain, so that in this respect a Hullian is necessarily a centralist. This locus is admitted by psychologists who have no scientific interest in it. But we cannot agree with the view that the centralism-peripheralism distinction is irrelevant. It is not, however, to be exhibited in a difference of data language, nor in the theorist's mere admission that the brain is somehow involved in the mediation of behavior! The difference lies in the physical locus assigned to *certain specified events characterized by the postulates*. And if a theorist makes assertions assigning such a locus, these assertions are just as much a part of the "real theory" as any of his other assertions. A methodological proposal has been made (e.g., 47) to define behavioral theory very restrictively, as by confining the *content* of theory to the

mathematical equations that occur in it. This is well and good as an expression of an interest, but somewhat arbitrary when used as a basis of exclusion against all other theoretical interests. At the very least, the theoretical meaning of learning constructs has to be given by a certain amount of non-mathematical context—we have at least to be informed that one of Hull's growth-functions refers to the change in habit strength rather than the growth of forelimbs. So far as the logic of science is concerned, we are surely not forbidden to introduce assertions of physical locus, or membership of an occurrence in a class specifiable in the language of another science, or actual identity of constructs inferred from originally unrelated empirical studies. If a theorist asserts, say, "The physical locus of the event designated  $r_G$  in theory  $T$  is in the periphery: that is,  $r_G$  is an effector-activity," Kendler's proposal seems to us tantamount to informing the theorist that this utterance is not part of his *real* theory. This is as if one were to inform the author of a genetics text that he had better delete all propositions identifying his statistical genetic constructs with certain physical loci on the visible rods called chromosomes, because this identification is not part of "real theory." Isn't this a rather arbitrary definition of what the theory consists of? Certainly it cannot be derived from any tenets of the general logical empiricism shared by all of us as scientists. The plain fact is that a scientist's theory concerning a specified empirical domain consists of the entire set of empirically meaningful sentences he asserts regarding it. One may not think it was clever of him to assert some of them; one may have little or no interest in some of them, such as those which identify constructs with those of allied sciences; one may accept most of the theory but reject or suspend judgment regarding the locus- or identity-assertions. All these attitudes are available to a critic such as Kendler; but what is *not* available to him is any sort of rule which will *exclude* from the theory some of the empirically meaningful propositions found in it. Nor are we here indulging in any confusion between Reichenbach's two contexts of discovery and justification (75, pp. 6-7). It is not a question of

whether the personal imagery of Hull or Tolman during their creative hours constitutes part of the theory. Of course it does not. But if what finally emerges is a set of significant sentences, they are all part of the theory whether some of us are interested in them or not.

Suppose, for example, that an inference postulate is not needed, but is derived as a theorem by some use of, say,  $r_G$ . The obvious Tolman retort is, "Very good. Capital. Only, I don't believe it." Now, just what is it the Tolmanite doesn't believe? One might say that, having admitted the derivability of the inference postulate from postulates not of that form, but which involve throughout the strengthening of *S-R* bonds, there is nothing left to be "denied." We cannot admit this. The rat is a physical mechanism, with parts whose structure and function are the subject-matter of several non-behavior sciences. There is, therefore, nothing "metaphysical" or "transcendent" about the question, "are the expectancies of the inference-postulates in fact mediated by chains of habits, the response terms of which are effector-event-classes?" The degree of confirmation of such a hypothesis *by behavior studies alone* cannot, perhaps, be very high. But this merely shows that a dogged "molar behaviorist" can ask more intelligible questions than he can easily answer by relying wholly upon his preferred behavioral methods. One need not be interested in Tolman's question; it is quite likely that such a question is currently not a profitable one. But neither of these is tantamount to a denial of its empirical meaningfulness.

So, our Tolmanite doesn't believe that the inference occurs via any such effector event as, say,  $r_G$ . He points to experimental data that seem to necessitate an inference postulate (assuming for the present discussion that he can do this and make it stick). The non-expectancy theorist grants him the postulate at the "first level" below behavior laws, but derives it as a theorem. The terms which occur in this derivation include not only the stimulus and response classes designated in the postulate, but certain other constructs such as  $r_G$ . In a concrete use of the postulate (theorem), this means the invocation of chewing-

movements, sense-organ-adjustments and the like. One can spend considerable time trying to invent experimental designs which will raise or lower the confirmation of such concrete applications of the "implicit response" type of construct, but the task is not easy. This is partly because, for all of the emphasis on quantification by non-expectancy theorists and especially by Hullians, the fractional goal response is at present almost wholly unquantified, so that it can be applied to almost any experimental outcome as an "out" with about as much abandon as Tolman can invoke attention, emphasis or the map's "strip-width" when the going gets hard. There are, however, some designs which, if they yielded positive results would render a peripheral mediator so unlikely that a centralist would be justified in requesting more direct (physiological) confirmation.

Another complaint a neo-Tolmanite would make is the total neglect of "perceptual organization" in our formulation. We have spoken of the *elicitor* as what gives rise to the *expectant*; whereas many centralists would say that in so doing we had skipped over what is the first, big step in learning—the transition from the elicitor to some central state consequent upon proximal stimulation, the configural properties of which are causally antecedent to the arousal of an expectant. We have avoided some aspects of this problem by emphasizing (*a*) the class-character of the elicitor and (*b*) the unlimited possibilities for configural specification of the (physical) stimulus side. But we are aware that a staunch centralist with perceptual interests will not feel these are sufficient. This problem was not discussed at any length at the Dartmouth Conference, chiefly because of the conferees' stronger interests in other issues. What little discussion of the topic took place in our sessions indicated considerable disagreement among us as to the importance and even as to the meaningfulness of certain perceptual distinctions. Consequently, we have nothing to say here except that we are fully aware of the omission of such considerations from our formalization proposals. The interposition of another intervening state, say, the *elicitant*, would not necessarily require profound altera-

tions in the remaining axioms. As we have chosen a quasi-stimulus language to characterize the expectate, so a quasi-stimulus language could be chosen for the elicitant's characterization. Postulates would then be needed to indicate the lawful relationships assumed to hold between the proximal elicitor and the (central) elicitant. Even if the remaining axioms remained unchanged, except for the substitution of *elicitant* for *elicitor* in referring to the sequence  $S_1 \rightarrow R_1 \rightarrow S_2$ , such a change would obviously have a considerable effect on the laws relating observables. Thus, number of exposures, even if the experimental procedure guaranteed adequate proximal representation of the elicitor, would take on a new meaning in the working of the whole theoretical system. If genuine discontinuities in the acquisition-function for *elicitants* were allowed for (e.g., sudden alterations in figure-ground relationships) the form of many derivative laws would be radically altered. It is our belief that the initial efforts at confirmation of expectancy assumptions ought *not* to be complicated by the use of experimental designs rich in perceptual-reorganizing possibilities. We do not think this reflects merely a bias for the response side, although it may in part. But there is present, regardless of one's bias, a peculiar asymmetry in the confirmation-relations of perceptual and expectancy postulates. The changes which occur in the animal's perceptual field as a function of exposures to the stimulus situation are inferred (or, constructed) from behavior-changes. Unless it is assumed that no quantitatively important changes are concurrently taking place in his expectancies (of which the perception is elicitant), we have two processes occurring at once, and their effect on the behavior is serial and cumulative. The components ( $y_1, y_2, \dots, y_m$ ) characterizing the perceptual field are related via some functions  $g_1, g_2, \dots, g_m$  to the components of the stimulus side; and the behavior in turn is functionally dependent  $f(y_1, y_2, \dots, y_m)$  upon the perceptions. In the case of human verbal reports made during perception experiments, we ordinarily operate on the plausible assumption that the values of the  $f$ -functions are very near their asymptotes (i.e., saying "red" when one

*perceives* red has been thoroughly overlearned in the pre-experimental history). Hence, we can determine the form of the *g*-functions by studying the relation between verbal responses and the stimulus variables. But the corresponding inference from response to perceptual field is obviously much more dangerous in the case of infra-human organisms.

The other horn of the dilemma is much less serious. As Spence has emphasized in his papers on the discontinuity controversy (93, 94, 95), we ought to choose stimuli of such a nature that the *g*-functions are of negligible importance. We can, of course, never be sure of this either, but we can usually be surer of it than we can of the expectancy (or habit) strengths. If an experimenter employs a goal-box, the floor and all walls of which are white, it is hard to doubt that the rat “perceives” its whiteness provided he has his eyes open. The curve of acquisition of an expectancy having this as the expectandum can then be investigated with some assurance, despite a possible lack of “complete psychological correlation” between the perception and the proximal stimulus. Once having such a law, we can investigate expectanda which are “perceptually” more interesting.

If, on the other hand, we should try to confirm the expectancy-acquisition postulates by utilizing—e.g., form discrimination—a negative result, or a quantitative outcome of a certain order would be contaminated in indeterminate ways and amounts by the well-known slowness of rats in making discriminations on the basis of form. If the “appropriate” behavior failed to appear following a latent learning procedure, it might merely indicate that the rat “had barely started to distinguish the triangle *as such*,” and hence prove nothing as to the expectancy postulates proper. Such interpretive ambiguities are in principle inevitable in the early stages of experimentation upon multi-variable systems, as we pointed out in our preliminary discussion of quantification. But there is little justification for inviting them by one’s choice of design and apparatus.

## SUMMARY AND CONCLUSIONS

We find it hard to bring together the preceding in a summary fashion. In its main outlines, we have seen no reason to criticize Tolman's approach as a *kind* of behavioral theory. Like its other critics, we wonder at its lack of even a minimal amount of formalization over the 20 years of controversy and experiment. It is hard to say to what extent this informality accounts for its viability under some very determined attacks. The question of its "factual adequacy" is a difficult one for this very reason, but we have concluded that within the limits imposed by its predictive vagueness, it cannot be said to stand refuted by the body of experimental evidence most clearly relevant—the latent learning studies. These findings seemed sufficiently encouraging to lead us to a preliminary attempt at formalization. The essential feature of that formalization is the *expectancy*, a cognitive unit sharing with *S-R* theory the explicit notational reference to response, but differing from *S-R* theory in its equally explicit reference to the expected consequences of responding. It is proposed that development of expectancy theory in this direction will be more fruitful than in the direction of "maps" and "perceptions" currently favored by Tolman himself.

At the risk of triviality, we cannot close without mentioning the contribution of the theory in generating informative experiments. Even if *S-R* theory should turn out to be, ultimately, the "whole truth" of the matter, no one would deny that the form of such an *S-R* theory will have been profoundly molded by the character of the opposition and the facts of learning discovered because of it. In closing, we would like to remind the reader of the words penned by R. M. Elliott in his editorial introduction to *Purposive Behavior in Animals and Men*:

Professor Tolman's argument may be ignored in some quarters; it will certainly be amplified by himself and others as new and crucial research data come to light; it will not, I think, be radically revised, that is "disproved"; and it will never be discredited, that is, shown to be either fictitious or unnecessary. Behaviorism of this sort has come of age (128, pp. viii-ix).

## Bibliography

1. Bendig, A. W. Latent learning in a water maze. *J. exp. Psychol.*, 1952, 43, 134-137.
2. Blodgett, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. publ. Psychol.*, 1929, 4, 113-134.
3. Broad, C. D. The "nature" of a continuant. In H. Feigl and W. Sellars (Eds.), *Readings in philosophical analysis*. New York: Appleton-Century-Crofts, 1949. Pp. 472-481.
4. Brown, J. S., & Farber, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, 48, 465-495.
5. Buxton, C. E. Latent learning and the goal gradient hypothesis. *Contr. psychol. Theor.*, 1940, 2, #2.
6. Carlin, Jean E. Drive stimulus generalization. Unpublished M.A. thesis, University of Minnesota, 1952.
7. Carnap, R. Testability and meaning. *Phil. Sci.*, 1937, 4, 1-40.
8. Carnap, R. The two concepts of probability. *Phil. phenomenol. Res.*, 1945, 5, 513-532.
9. Carnap, R. Empiricism, semantics, and ontology. *Rev. int. de Phil.*, 1950, 4, 20-40.
10. Carr, H. Teaching and learning. *J. genet. Psychol.*, 1930, 37, 189-218.
11. Chisholm, R. M. Intentionality and the theory of signs. *Phil. Stud.*, 1952, 3, 56-63.
12. Christie, R. Experimental naivete and experiential naivete. *Psychol. Bull.*, 1951, 48, 327-339.
13. Christie, R. The role of drive discrimination in learning under irrelevant motivation. *J. exp. Psychol.*, 1951, 42, 13-19.
14. Christie, R. The effect of some early experiences in the latent learning of adult rats. *J. comp. physiol. Psychol.*, 1952, 43, 281-288.
15. Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, 55, 467-517.
16. Crespi, L. P. Amount of reinforcement and level of reinforcement. *Psychol. Rev.*, 1944, 51, 341-357.
17. Daub, C. T. The effect of doors on latent learning. *J. comp. Psychol.*, 1933, 15, 49-58.

18. Dennis, W. A comparison of the rat's first and second explorations of a maze unit. *Amer. J. Psychol.*, 1935, 47, 488-490.
19. Dennis, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effects. *J. comp. Psychol.*, 1939, 28, 305-312.
20. Dennis, W., & Sollenberger, R. T. Negative adaption in the maze exploration of albino rats. *J. comp. Psychol.*, 1934, 18, 197-206.
21. Denny, M. R., & Davis, R. H. A test of latent learning for a non-goal significate. *J. comp. physiol. Psychol.*, 1951, 44, 590-595.
22. Diesenroth, C. F., & Spence, K. W. An investigation of latent learning in the white rat. *Psychol. Bull.*, 1941, 38, 706. (Abstract)
23. Elliott, M. H. The effect of appropriateness of rewards and of complex incentives on maze performance. *Univ. Calif. publ. Psychol.*, 1929, 4, 91-98.
24. Estes, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, 57, No. 3 (Whole No. 263).
25. Estes, W. K. Effects of competing reactions on the conditioning curve for bar-pressing. *J. exp. Psychol.*, 1950, 40, 200-205.
26. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, 57, 94-107.
27. Estes, W. K., & Skinner, B. F. Some quantitative properties of anxiety. *J. exp. Psychol.*, 1941, 29, 390-400.
28. Fehrer, Elizabeth. Latent learning in the sophisticated rat. *J. exp. Psychol.*, 1951, 42, 409-416.
29. Finan, J. L. Quantitative studies in motivation. I. Strength of conditioning in rats under varying degrees of hunger. *J. comp. Psychol.*, 1940, 29, 119-134.
30. Gilchrist, J. C. Characteristics of latent and reinforcement learning as a function of time. *J. comp. physiol. Psychol.*, 1952, 45, 198-203.
31. Gleitman, H. Studies in motivation and learning: II. Thirsty rats trained in maze with food but not water; then run hungry. *J. exp. Psychol.*, 1950, 40, 169-174.
32. Grice, G. R. An experimental test of the expectation theory of learning. *J. comp. physiol. Psychol.*, 1940, 41, 137-143.
33. Haney, G. W., The effect of familiarity on maze performance of albino rats. *Univ. Calif. publ. Psychol.*, 1931, 4, 319-333.
34. Hempel, C. G. Problems and changes in the empiricist criterion of meaning. *Rev. int. de Phil.*, 1950, 4, 41-63.
35. Herb, F. H. Latent learning—non-reward followed by food in blinds. *J. comp. Psychol.*, 1940, 29, 247-255.

36. Heron, W. T. Internal stimuli and learning. *J. comp. physiol. Psychol.*, 1949, 42, 486-492.
37. Hilgard, E. R. *Theories of learning*. New York: Appleton-Century-Crofts, 1948.
38. Hull, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
39. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, 57, 173-180.
40. Hull, C. L. *A behavior system*. New Haven: Yale University Press, 1952.
41. Iwahara, S., & Marx, M. Cognitive transfer in discrimination learning. *Amer. Psychologist*, 1950, 5, 479. (Reported by title only.)
42. Karn, H. W., & Porter, H. M., Jr. The effects of certain pre-training procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, 36, 461-469.
43. Keller, F. S., & Schoenfeld, W. N. *Principles of Psychology*. New York: Appleton-Century-Crofts, 1950.
44. Kendler, H. H. Drive interaction: II. Experimental analysis of the role of drive in learning theory. *J. exp. Psychol.*, 1945, 35, 188-198.
45. Kendler, H. H. A comparison of learning under motivated and satiated conditions in the white rat. *J. exp. Psychol.*, 1947, 37, 545-549.
46. Kendler, H. H. An investigation of latent learning in a T-maze. *J. comp. physiol. Psychol.*, 1947, 40, 265-270.
47. Kendler, H. H. "What is learned?"—a theoretical blind alley. *Psychol. Rev.*, 1952, 59, 269-277.
48. Kendler, H. H., & Kanner, J. H. A further test of the ability of rats to learn the location of food when motivated by thirst. *J. exp. Psychol.*, 1950, 40, 762-765.
49. Kendler, H. H., & Mencher, H. C. The ability of rats to learn the location of food when motivated by thirst—an experimental reply to Leeper. *J. exp. Psychol.*, 1948, 38, 82-88.
50. Krechevsky, I. "Hypotheses" versus "chance" in the presolution period in sensory discrimination-learning. *Univ. Calif. publ. Psychol.*, 1932, 6, 27-44.
51. Lashley, K. S. A simple maze: with data on the relation of the distribution of practice to rate of learning. *Psychobiol.*, 1918, 1, 353-367.
52. Leeper, R. W. The role of motivation in learning: a study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, 46, 3-40.
53. Leeper, R. W. The experiments by Spence and Lippitt and by Kendler on the

- sign-gestalt theory of learning. *J. exp. Psychol.*, 1948, 38, 102-106.
54. Littman, R. A. Latent learning in a T-maze after two degrees of training. *J. comp. physiol. Psychol.*, 1950, 43, 135-147.
  55. Littman, R. A., & Rosen, E. Molar and molecular. *Psychol. Rev.*, 1950, 57, 58-65.
  56. MacCorquodale, K., & Meehl, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, 55, 95-107.
  57. MacCorquodale, K., & Meehl, P. E. "Cognitive" learning in the absence of competition of incentives. *J. comp. physiol. Psychol.*, 1949, 42, 383-390.
  58. MacCorquodale, K., & Meehl, P. E. On the elimination of cul-entries without obvious reinforcement. *J. comp. physiol. Psychol.*, 1951, 44, 367-371.
  59. MacCorquodale, K., & Meehl, P. E. Preliminary suggestions as to a formalization of expectancy theory. *Psychol. Rev.*, 1953, 60, 55-63.
  60. MacDuff, Mary A. The effect on retention of varying degrees of motivation during learning in rats. *J. comp. physiol. Psychol.*, 1943, 33, 369-386.
  61. Maltzman, I. M. An experimental study of learning under an irrelevant need. *J. exp. Psychol.*, 1950, 40, 788-793.
  62. Meehl, P. E. An examination of the treatment of stimulus patterning in Prof. Hull's Principles of Behavior. *Psychol. Rev.*, 1945, 52, 324-332.
  63. Meehl, P. E. On the circularity of the law of effect. *Psychol. Bull.*, 1950, 47, 52-75.
  64. Meehl, P. E., & MacCorquodale, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, 41, 372-396.
  65. Meehl, P. E., & MacCorquodale, K. A failure to find the Blodgett effect, and some secondary observations on drive conditioning. *J. comp. physiol. Psychol.*, 1951, 44, 178-183.
  66. Meehl, P. E., & MacCorquodale, K. Some methodological comments concerning expectancy theory. *Psychol. Rev.*, 1951, 58, 230-233.
  67. Meehl, P. E., & MacCorquodale, K. Drive conditioning as a factor in latent learning. *J. exp. Psychol.*, 1953, 43, 20-24.
  68. Miller, N., & Dollard, J. *Social learning and imitation*. New Haven: Yale University Press, 1941.
  69. Montgomery, K. C. "Spontaneous alternation" as a function of time between trials and amount of work. *J. exp. Psychol.*, 1951, 42, 82-93.
  70. Montgomery, K. C. The relation between exploratory behavior and spontaneous alternation in the white rat. *J. comp. physiol. Psychol.*, 1951, 44, 582-589.

71. Montgomery, K. C. Exploratory behavior and its relation to spontaneous alternation in a series of maze exposures. *J. comp. physiol. Psychol.*, 1952, 45, 50-57.
72. Muenzinger, K. F. Plasticity and mechanization of the problem box habit in guinea pigs. *J. comp. Psychol.*, 1928, 8, 45-69.
73. Muenzinger, K. F., Koerner, L., & Ireys, E. Variability of an habitual movement in guinea pigs. *J. comp. Psychol.*, 1929, 9, 425-436.
74. Nissen, H. W. Description of the learned response in discrimination behavior. *Psychol. Rev.*, 1950, 57, 121-131.
75. Reichenbach, H. *Experience and prediction*. Chicago: Univ. Chicago Press, 1938.
76. Reynolds, B. A repetition of the Blodgett experiment on "latent learning." *J. exp. Psychol.*, 1945, 35, 504-516.
77. Reynolds, B. The relationship between the strength of a habit and the degree of drive present during acquisition. *J. exp. Psychol.*, 1949, 39, 296-305.
78. Ritchie, B. F. Studies in spatial learning. III. Two paths to the same location and two paths to two different locations. *J. exp. Psychol.*, 1947, 37, 25-38.
79. Ritchie, B. F. Studies in spatial learning. VI. Place orientation and direction orientation. *J. exp. Psychol.*, 1948, 38, 659-669.
80. Ritchie, B. F., Aeschliman, B., & PEIRCE, P. Studies in spatial learning. VIII. Place performance and the acquisition of place dispositions. *J. comp. physiol. Psychol.*, 1950, 43, 73-85.
81. Ritchie, B. F., Hay, Alice, & Hare, Rachel. Studies in spatial learning: IX. A dispositional analysis of response performance. *J. comp. physiol. Psychol.*, 1951, 44, 442-449.
82. Sellars, W. S. Concepts as involving laws and inconceivable without them. *Phil. of Sci.*, 1948, 15, 287-315.
83. Sellars, W. S. Mind, meaning, and behavior. *Philos. Studies*, 1952, 3, 83-95.
84. Seward, J. P. A theoretical derivation of latent learning. *Psychol. Rev.*, 1947, 54, 83-98.
85. Seward, J. P. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, 39, 177-186.
86. Seward, J. P., Datel, W. E., & Levy, N. Tests of two hypotheses of latent learning. *J. exp. Psychol.*, 1952, 43, 274-280.
87. Seward, J. P., Levy, N., & Handlon, J. P., Jr. Incidental learning in the rat. *J. comp. physiol. Psychol.*, 1950, 43, 240-251.

88. Shaw, M. E., & Waters, R. H. An experimental test of latent learning in a relatively free-choice situation. *J. genet. Psychol.*, 1950, 77, 283-292.
89. Simmons, Rietta. The relative effectiveness of certain incentives in animal learning. *Comp. Psychol. Monogr.*, 1924, 2, 1-79.
90. Skinner, B. F. *The behavior of organisms*. New York: Appleton-Century-Crofts, 1938.
91. Skinner, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, 57, 193-216.
92. Solomon, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, 55, 1-40.
93. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, 44, 430-444.
94. Spence, K. W. Continuous versus non-continuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, 47, 271-288.
95. Spence, K. W. An experimental test of the continuity and non-continuity theories of discrimination learning. *J. exp. Psychol.*, 1945, 35, 253-266.
96. Spence, K. W. The methods and postulates of "behaviorism." *Psychol. Rev.*, 1948, 55, 67-78.
97. Spence, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: John Wiley and Sons, 1951.
98. Spence, K. W., Bergmann, G., & Lippitt, R. A study of simple learning under irrelevant motivational-reward conditions. *J. exp. Psychol.*, 1950, 40, 539-551.
99. Spence, K. W., & Kendler, H. H. The speculations of Leeper with respect to the Iowa tests of the sign-gestalt theory of learning. *J. exp. Psychol.*, 1948, 38, 106-109.
100. Spence, K. W., & Lippitt, R. "Latent" learning of a simple maze problem with relevant needs satiated. *Psychol. Bull.*, 1940, 37, 429. (Abstract)
101. Spence, K. W., & Lippitt, R. An experimental test of the sign-gestalt theory of trial and error learning. *J. exp. Psychol.*, 1946, 36, 491-502.
102. Strange, J. R. Latent learning under conditions of high motivation. *J. comp. physiol. Psychol.*, 1950, 43, 194-197.
103. Strassbitrger, R. C. Resistance to extinction of a conditioned operant as related to drive level at reinforcement. *J. exp. Psychol.*, 1950, 40, 473-487.
104. Szymanski, J. S. Versuch liber die Wirkung der Faktoren, die als Antrieb zum Erlernen einer Handlung dienen Konnen. *Pflüg. arch. ges. Physiol.*, 1918,

- 171, 374-385.
105. Tarski, A. *Introduction to logic*. New York: Oxford University Press, 1941.
  106. Thistlethwaite, D. L. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, 48, 97-129.
  107. Thistlethwaite, D. L. An experimental test of a reinforcement interpretation of latent learning. *J. comp. physiol. Psychol.*, 1951, 44, 431-441.
  108. Thistlethwaite, D. L. Conditions of irrelevant incentive learning. *J. comp. physiol. Psychol.*, 1952, 45, 517-525.
  109. Tinklepaugh, O. L. An experimental study of representative factors in monkeys. *J. comp. Psychol.*, 1928, 8, 197-236.
  110. Tolman, E. C. More concerning the temporal relations of meaning and imagery. *Psychol. Rev.*, 1917, 24, 114-138.
  111. Tolman, E. C. Retroactive inhibition as affected by conditions of learning. *Psychol. Monogr.*, 1918, 25, No. 107, 50 pp.
  112. Tolman, E. C. Nerve process and cognition. *Psychol. Rev.*, 1918, 25, 423-442.
  113. Tolman, E. C. Instinct and purpose. *Psychol. Rev.*, 1920, 27, 217-233.
  114. Tolman, E. C. A new formula for behaviorism. *Psychol. Rev.*, 1922, 29, 44-53.
  115. Tolman, E. C. Can instincts be given up in psychology? *J. abn. Psychol.*, 1922, 17, 139-152.
  116. Tolman, E. C. The nature of instinct. *Psychol. Bull.*, 1923, 20, 200-216.
  117. Tolman, E. C. A behavioristic account of the emotions. *Psychol. Rev.*, 1923, 30, 217-227.
  118. Tolman, E. C. The effects of underlearning upon long- and short-time retentions. *J. exp. Psychol.*, 1923, 6, 466-474.
  119. Tolman, E. C. The inheritance of maze-learning ability in rats. *J. comp. Psychol.*, 1924, 4, 1-18.
  120. Tolman, E. C. Behaviorism and purpose. *J. Phil.*, 1925, 22, 36-41.
  121. Tolman, E. C. Purpose and cognition: the determiners of animal learning. *Psychol. Rev.*, 1925, 32, 285-297.
  122. Tolman, E. C. The nature of the fundamental drives. *J. abn. soc. Psychol.*, 1926, 20, 349-358.
  123. Tolman, E. C. A behavioristic theory of ideas. *Psychol. Rev.*, 1926, 33, 352-369.
  124. Tolman, E. C. Habit formation and higher mental processes in animals. *Psychol. Bull.*, 1927, 24, 1-35; 1928, 25, 24-53.
  125. Tolman, E. C. A behaviorist's definition of consciousness. *Psychol. Rev.*, 1927,

- 34, 433-439.
126. Tolman, E. C. Purposive behavior. *Psychol. Rev.*, 1928, 35, 524-530.
  127. Tolman, E. C. Maze performance a function of motivation and of reward as well as of knowledge of the maze paths. *J. gen. Psychol.*, 1930, 4, 338-342.
  128. Tolman, E. C. *Purposive behavior in animals and men*. New York: Appleton-Century, 1932.
  129. Tolman, E. C. Lewin's concept of vectors. *J. gen. Psychol.*, 1932, 7, 3-15.
  130. Tolman, E. C. Sign-gestalt or conditioned reflex? *Psychol. Rev.* 1933, 40, 246-255.
  131. Tolman, E. C. The law of effect. A reply to Dr. Goodenough. *J. exp. Psychol.*, 1933, 16, 463-470.
  132. Tolman, E. C. Backward elimination of errors in two successive discrimination habits. *Univ. Calif. publ. Psychol.*, 1934, 6, 145-152.
  133. Tolman, E. C. Theories of learning. In F. A. Moss, *Comparative Psychology*. New York: Prentice-Hall, 1934. Pp. 367-408.
  134. Tolman, E. C. Psychology versus immediate experience. *Philos. Science*, 1935, 2, 356-380.
  135. Tolman, E. C. Operational behaviorism and current trends in psychology. *Proc. 25th Anniv. Inauguration Graduate Studies*. Los Angeles: University of Southern California, 1936, 89-103.
  136. Tolman, E. C. Connectionism: wants, interests, and attitudes. *Character and Personality*, 1936, 4, 245-253.
  137. Tolman, E. C. Operational behaviorism and current trends in psychology. *Proc. 25th Anniv. Inauguration Graduate Studies*. Los Angeles: University of Southern California, 1936, 89-103.
  138. Tolman, E. C. Demands and conflicts. *Psychol. Rev.*, 1937, 44, 158-169.
  139. Tolman, E. C. The acquisition of string-pulling by rats—conditioned response or sign-gestalt? *Psychol. Rev.*, 1937, 44, 195-211.
  140. Tolman, E. C. An operational analysis of "Demands." *Erkenntnis*, 1937, 6, 383-392.
  141. Tolman, E. C. The determiners of behavior at a choice point. *Psychol. Rev.*, 1938, 45, 1-41.
  142. Tolman, E. C. A reply to Professor Guthrie. *Psychol. Rev.*, 1938, 45, 163-164.
  143. Tolman, E. C. The law of effect. *Psychol. Rev.*, 1938, 45, 200-203.
  144. Tolman, E. C. Prediction of vicarious trial and error by means of the schematic sawbug. *Psychol. Rev.*, 1939, 46, 318-336.

145. Tolman, E. C. Spatial angle and vicarious trial and error. *J. comp. Psychol.*, 1940, 30, 129-135.
146. Tolman, E. C. Motivation, learning, and adjustment. *Proc. Amer. Philos. Soc.*, 1941, 84, 543-563.
147. Tolman, E. C. Discrimination vs. learning and the schematic sowbug. *Psychol. Rev.*, 1941, 48, 367-382.
148. Tolman, E. C. A drive-conversion diagram. *Psychol. Rev.*, 1943, 50, 503-513.
149. Tolman, E. C. A stimulus-expectancy need-cathexis psychology. *Science*, 1945, 101, 160-166.
150. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, 55, 189-208.
151. Tolman, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, 56, 144-155.
152. Tolman, E. C. Discussion: Interrelationships between perception and personality. *J. Personal.*, 1949, 18, 48-50.
153. Tolman, E. C. The psychology of social learning. *J. soc. Issues*, 1949, 5, Supplement No. 3, 5-18.
154. Tolman, E. C. The nature and functioning of wants. *Psychol. Rev.*, 1949, 56, 357-369.
155. Tolman, E. C., & Brunswik, E. The organism and the causal texture of the environment. *Psychol. Rev.*, 1935, 42, 43-77.
156. Tolman, E. C., & Davis, F. C. A note on the correlations between two mazes. *J. comp. Psychol.*, 1924, 4, 125-135.
157. Tolman, E. C., & Geier, F. M. Goal distance and restless activity. I: The goal gradient of restless activity. *J. comp. Psychol.*, 1943, 35, 197-204.
158. Tolman, E. C., Geier, F. M., & Levin, M. Individual differences in emotionality, hypothesis formation, vicarious trial and error, and visual discrimination learning in rats. *Comp. psychol. Monogr.*, 1941, 17, No. 3. 20 pp.
159. Tolman, E. C., & Gleitman, H. Studies in learning and motivation: I. Equal reinforcements in both end-boxes, followed by shock in one end-box. *J. exp. Psychol.*, 1949, 39, 810-819.
160. Tolman, E. C., & Gleitman, H. Studies in spatial learning: VII. Place and response learning under different degrees of motivation. *J. exp. Psychol.*, 1949, 39, 653-659.
161. Tolman, E. C., Hall, C. S., & Brettnall, E. P. A disproof of the law of effect and a substitution of the laws of emphasis, motivation, and disruption. *J. exp. Psychol.*, 1932, 15, 601-614.

162. Tolman, E. C., & Honzik, C. H. Introduction and removal of reward, and maze performance of rats. *Univ. Calif. publ. Psychol.*, 1930, 4, 257- 275.
163. Tolman, E. C., & Honzik, C. H. The perception of spatial relations by the rat: a type of response not easily explained by conditioning. *J. comp. Psychol.*, 1936, 22, 287-318.
164. Tolman, E. C., & Honzik, C. H. The action of punishment in accelerating learning. *J. comp. Psychol.*, 1938, 26, 187-200.
165. Tolman, E. C., & Honzik, C. H. "Insight" in rats. *Univ. Calif. publ. Psychol.*, 1930, 4, 215-232.
166. Tolman, E. C., & Honzik, C. H. Degrees of hunger, reward and non-reward, and maze learning in rats. *Univ. Calif. publ. Psychol.*, 1930, 4, 241-257.
167. Tolman, E. C., & Honzik, C. H. Introduction and removal of reward, and maze performance in rats. *Univ. Calif. publ. Psychol.*, 1930, 4, 257-275.
168. Tolman, E. C., & Honzik, C. H., & Robinson, E. W. The effect of degrees of hunger upon the order of elimination of long and short blinds. *Univ. Calif. publ. Psychol.*, 1930, 4, 189-202.
169. Tolman, E. C., & Horowitz, J. A reply to Mr. Koffka. *Psychol. Bull.*, 1933, 30, 459-465.
170. Tolman, E. C., & Krechevsky, I. Means-end-readiness and hypothesis. A contribution to comparative psychology. *Psychol. Rev.*, 1933, 40, 60-70.
171. Tolman, E. C., & Minium, E. VTE in rats: Overlearning and difficulty of discrimination. *J. comp. Psychol.*, 1942, 34, 301-306.
172. Tolman, E. C., & Ritchie, B. F. Correlation between VTE's on a maze and on a visual discrimination apparatus. *J. comp. Psychol.*, 1943, 36, 91-98.
173. Tolman, E. C., & Ritchie, B. F., & Kalish, D. Studies in spatial learning. I. Orientation and the short-cut. *J. exp. Psychol.*, 1946, 36, 13-24.
174. Tolman, E. C., & Ritchie, B. F., & Kalish, D. Studies in spatial learning. II. Place learning vs. response learning. *J. exp. Psychol.*, 1946, 36, 221-229.
175. Tolman, E. C., & Ritchie, B. F., & Kalish, D. Studies in spatial learning. IV. The transfer of place learning to other starting paths. *J. exp. Psychol.*, 1947, 37, 39-47.
176. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial learning. V. Response learning vs. place learning by the non-correction method. *J. exp. Psychol.*, 1947, 37, 285-292.
177. Tolman, E. C., & Sams, C. F. Time discrimination in white rats. *J. comp. Psychol.*, 1925, 5, 255-263.
178. Tolman, E. C., & White, A. E. A note on the elimination of short and long blind

- alleys. *J. comp. Psychol.*, 1923, 3, 327-332.
179. Walker, E. L. The acquisition of a response to food under conditions of food satiation. *Amer. Psychologist*, 1948, 3, 239. (Reported by title only.)
180. Walker, E. L. Drive specificity and learning. *J. exp. Psychol.*, 1948, 38, 39-49.
181. Walker, E. L. The demonstration of learning acquired under a strong irrelevant drive previously masked by a primarily reinforced response. *Amer. Psychologist*, 1950, 5, 479. (Reported by title only.)
182. Walker, E. L. Drive specificity and learning: Demonstration of a response tendency acquired under a strong irrelevant drive. *J. comp. physiol. Psychol.*, 1951, 44, 596-603.
183. Walker, E. L., Knottter, M. C., & DEVALOIS, R. L. Drive specificity and learning: The acquisition of a spatial response to food under conditions of water deprivation and food satiation. *J. exp. Psychol.*, 1950, 40, 161-168.
184. Wallace, S. R., Jr., Blackwell, M. G., Jr., & Jenkins, G. Pre-reward and post-reward performance in the "latent learning" of an elevated maze. *Psychol. Bull.*, 1941, 38, 694. (Abstract)
185. White, R. K. The case for the Tolman-Lewin interpretation of learning. *Psychol. Rev.*, 1943, 50, 157-186.
186. Williams, K. A. The reward value of a conditioned stimulus. *Univ. Calif. publ. Psychol.*, 1929, 4, 31-55.
187. Wingfield, R. C., & Dennis, W. The dependence of the white rat's choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, 1934, 18, 135-148.