VOL. 44, NO. 1 JANUARY, 1937

THE PSYCHOLOGICAL REVIEW

MIND, MECHANISM, AND ADAPTIVE BEHAVIOR [1]

BY CLARK L. HULL

Institute of Human Relations, Yale University

INTRODUCTION

Since the time of Charles Darwin it has become clear not only that living organisms have gradually evolved through immense periods of time, but that man is evolution's crowning achievement. It is equally clear that man's preëminence lies in his capacity for adaptive behavior. Because of the seemingly unique and remarkable nature of adaptive behavior, it has long been customary to attribute it to the action of a special agent or substance called 'mind.' Thus 'mind' as a hypothetical entity directing and controlling adaptive behavior attains biological status possessing survival value and, consequently, a 'place in nature.' But what is this mysterious thing called mind? By what principles does it operate? Are these principles many or are they few? Are they those of the ordinary physical world or are they of the nature of spiritual essences—of an entirely different order, the non-physical?

It will, perhaps, be most economical to begin our examination of this important problem by passing briefly in review some typical phenomena of adaptive behavior which have led to the assumption of a special psychic entity. Among these may be mentioned the following: When obstacles are encountered, organisms often persist in making the same incorrect attempt over and over again; they vary their re-

[1] Presidential Address delivered before the American Psychological Association, Hanover, New Hampshire, September 4, 1936.

The author is indebted to Professor Max Wertheimer for a critical reading of this paper.

actions spontaneously; they display anticipatory reactions antedating the biological emergencies to which the reactions are adaptive; they present the phenomena of disappointment and discouragement; they strive to attain states of affairs which are biologically advantageous; they transfer to new problem situations adaptive behavior acquired in situations which, objectively considered, are totally different. The behavior of organisms is purposive in that they strive for goals or values, and in so doing manifest intelligence or insight and a high degree of individual freedom from current coercion of the environment. Whatever may be the final conclusion as to the ultimate nature of these phenomena, their biological significance in terms of survival must be immense. The task of understanding and controlling them is surely worthy of the best coöperative efforts of the biological and social sciences.

The Controversy Regarding Adaptive Behavior Is Theoretical, Not Factual

Historically, two main views have been held as to the ultimate nature of adaptive behavior. The most widely accepted of these, at the present time, is also the most ancient; its roots lie far back in primitive animism. According to this view, the principles governing adaptive behavior are essentially non-physical, mental, or psychic. The second view, despite its austerity, has received a certain amount of favor among men of science. It assumes that adaptive behavior operates ultimately according to the principles of the physical world. In our consideration of these contrasting views, it will be convenient to begin with the latter.

The physical or mechanistic view of the nature of adaptive behavior can best be stated by quoting the beautiful presentation of the raindrop analogy written by the late Albert P. Weiss:

> We may best visualize the relationship between the responses that make up the so-called purposive behavior category by the raindrop analogy. We may start with the assumption that every drop of rain in some way or other gets to the

> ocean. . . . Anthropomorphizing this condition we may say that it is the *purpose* of every drop of rain to get to the ocean. Of course, this only means that virtually every drop *does* get there eventually. . . . Falling from the cloud it may strike the leaf of a tree, and drop from one leaf to another until it reaches the ground. From here it may pass under or on the surface of the soil to a rill, then to a brook, river, and finally to the sea. Each stage, each fall from one leaf to the next, may be designated as a *means* toward the final end, the sea, . . . Human behavior is merely a complication of the same factors.[2]

The nub of Weiss's statement lies in his concluding remark that adaptive behavior is merely a 'complication' of the same factors as those which are involved in the behavior of a drop of water finding its way from an inland cloud to the sea. Obviously, Weiss did not mean to say that the several forms of seeking and striving behavior characteristic of the higher organisms are brought about by the various compoundings of such processes as evaporation, condensation, splashing, and flowing. The context of the quotation shows that he meant that ultimately the complex forms of purposive behavior would be found to derive from the same *source* as those from which the raindrop phenomena are derived; *i.e.*, from the basic entities of theoretical physics, such as electrons and protons. He discusses these latter concepts explicitly and at length.

Passing to the more orthodox view, that adaptive behavior is essentially non-physical, or psychic, the words of A. S. Eddington may be taken as a point of departure. In his book, 'The nature of the physical world,'[3] Eddington remarks:

> Conceivably we might reach a human machine interacting by reflexes with its environment; but we cannot reach rational man morally responsible. [P. 343.] . . . In a world of æther and electrons we might perhaps encounter *nonsense;* we could not encounter *damned nonsense.*

[2] Albert P. Weiss, A theoretical basis of human behavior, Columbus, Ohio: R. G. Adams and Company, 1925, pp. 346–347.

[3] New York: The Macmillan Company, 1929, p. 345.

The significance of Eddington's statement centers around the word *reach*. From the present point of view, he seems to be saying that we cannot reach the highest forms of adaptive behavior, such as complex problem solution (rational behavior) and certain complex forms of social behavior involving the implicit verbal coercion of the behavior of the individual (moral behavior) if we start out merely with æther and electrons; we must begin with something non-physical, or psychic—presumably consciousness.

Thus the issue is joined. We are presented with the paradox of Eddington, the physicist, apparently insisting that the higher forms of behavior are at bottom non-physical, whereas Weiss, the psychologist, insists that they are fundamentally non-phychological!

But what, exactly, is the issue? Is it, for example, a difference as to an ordinary matter of observed fact? Do Eddington and those who share his view claim to have made certain observations which are in conflict with a corresponding set of observations supposed to have been made by Weiss and those with a mechanistic leaning? The dispute involves nothing of this nature. It is clear that the controversy is definitely a theoretical one. Eddington seems to be implying that we *can not* reach a sound theory of rational, purposive and moral behavior if we set out with nothing but æther and electrons. Weiss is saying, by implication, that a sound theory of such behavior *can* be reached by setting out with nothing but electrons and protons.

The Methodology of Scientific Theory Differentiated from That of Philosophical Speculation

Having located definitely in the field of theory the contrasted views represented in a general way by Weiss and Eddington, we face at once the critical question of whether the problem lies within the range of the operation of scientific methodology. If it does, what is that methodology? How is it to be applied to the question before us in a way which will avoid the interminable wrangles and philosophical futilities so long associated with the mind-body problem? It will be necessary to go into the matter of methodology rather thoroughly, in part because of its central importance for our present problem, but in part also because of the widespread misconceptions regarding it due to our early associations with philosophy. With the question of methodology clarified we shall return to Weiss and Eddington in the hope of demonstrating its concrete application.

The essential characteristics of a sound scientific theoretical system, as contrasted with ordinary philosophical speculation, may be briefly summarized under three heads:

1. A satisfactory scientific theory should begin with a set of explicitly stated postulates accompanied by specific or 'operational' definitions of the critical terms employed.

2. From these postulates there should be deduced by the most rigorous logic possible under the circumstances, a series of interlocking theorems covering the major concrete phenomena of the field in question.

3. The statements in the theorems should agree in detail with the observationally known facts of the discipline under consideration. If the theorems agree with the observed facts, the system is probably true; if they disagree, the system is false. If it is impossible to tell whether the theorems of a system agree with the facts or not, the system is neither true nor false; scientifically considered, it is meaningless.

Since concrete example is more illuminating and more convincing than abstract statement, there is reproduced below a small scientific theoretical system in which an attempt has been made to conform to the above principles. There may be found (p. 15 ff.) a number of definitions, which are followed (p. 16 ff.) by six postulates. The system concludes with a series of thirteen theorems (p. 17 ff.), each derived from the postulates by a process of reasoning analogous to that ordinarily employed in geometry.

At first sight the formal characteristics of scientific theory look very much like those of philosophical speculation and even of ordinary argumentation, from which philosophical speculation can scarcely be distinguished. At their best, both scientific theory and philosophical speculation set out from explicit postulates; both have definitions of critical terms; both have interlocking theorems derived by meticulous logic. Consider, for example, Spinoza's 'Ethic,' a philosophical work of the better sort. This has all of the above characteristics in almost exactly the same form as the miniature scientific system which is presented below. Where, then, lie the great difference and superiority of the scientific procedure?

The answer, while extending into many complex details, rests upon a single fundamental principle. The difference is that *in philosophical speculation there is no possibility of comparing a theorem with the results of direct observation.* An obvious example of this impossibility is seen in Spinoza's famous pantheistic theorem, Proposition XIV, from Part One of his 'Ethic':

> *Besides God no substance can be, nor can be conceived.*

It is difficult to imagine subjecting such a theorem as that to an observational test.

Consider, by way of contrast, a really scientific procedure, one carried out by Galileo at about the same time that Spinoza was writing. The Copernican hypothesis concerning the nature of the solar system was then in violent dispute. From this hypothesis, together with a few familiar principles concerning the behavior of light, it follows logically as a theorem that the planet Venus, like the moon, should show the crescent and all the other stages between the full and dark phases. Presumably led by this deduction, Galileo, with a telescope of his own construction, made the necessary observations on Venus and found the phases exactly as demanded by the theorem. Here we have the indispensable observational check demanded by science but lacking in philosophy.

But why, it will be asked, is it so imperative to have an observational check on the theorems of a system if the system is to merit serious consideration by scientists? To answer this question adequately it will be necessary to consider in a little detail the characteristics of postulates, the procedure in selecting them, and the methodology of their substantiation.

It is important to note at the outset that in scientific theory postulates tend to be of two kinds. First, there are postulates which are mere matters of fact; *i.e.*, they are matters of relatively simple and direct observation. Second, there are postulates which by their nature cannot conceivably be matters of direct observation. The classical investigation of Galileo just considered contains examples of both types. The principles of light and shadow upon which lunar and planetary phases depend are obviously matters of ordinary, everyday, direct terrestrial observation, and so represent postulates of the first type. On the other hand, the Copernican hypothesis as to the relative movements of the several components of the solar system is not susceptible to direct observation, and so represents postulates of the second type.

In scientific theory, owing to the continuous checking of theorems arrived at deductively against the results of direct observation, both types of postulates are constantly receiving *indirect* verification or refutation. Thus postulates capable of the direct approach are susceptible of two independent kinds of test, the direct and the indirect. But the continuous indirect test is of special importance for the postulates incapable of the direct approach. Were it not for this they would be subject to no observational verification at all, and scientific theory would in this respect have no more safeguard against erroneous basic assumptions than has philosphical speculation. Thus Galileo's brilliant observations of the phases of Venus not only gave the scientific world some new facts but, of far greater importance, they substantiated in a convincing, though indirect, manner the fundamental Copernican hypothesis.[4]

[4] Many persons have been puzzled by the paradox that in science a deduction frequently sets out with postulates which are by no means securely established, whereas in ordinary argumentation there is the greatest insistence upon the certainty of the premises upon which the argument is based. The explanation of this paradox lies largely in the difference of objective in the two cases. Argument ordinarily seeks to convince by a deductive procedure of something which under the circumstances is not directly observable; otherwise there would be no point in performing the deduction. It is clear that if the person to whom the argument is directed does not agree with the premises he will not agree with the conclusion and the whole procedure will

Whenever a theorem fails to check with the relevant facts, the postulates which gave rise to it must be ruthlessly revised until agreement is reached. If agreement cannot be attained, the system must be abandoned. In this constant revision there is a definite tendency to choose and formulate the postulates in a way which will make them yield the deductions desired. Such a procedure involves an obvious element of circularity. This is particularly the case where the system is small and where the postulates are purely symbolic constructs or inventions and therefore not subject to direct investigation. Even so, the choice of postulates to fit the facts is methodologically legitimate and, upon the whole, desirable. One important reason for this is that a postulate or hypothesis so arrived at may lead to a *direct*, experimental confirmation in case it is capable of the direct approach.[5] In such an event, of course, all circularity disappears.

But if the system is truly scientific in nature, the circularity just considered is only a temporary phase even when one or more of the postulates are insusceptible to direct investigation. It is precisely in this connection that scientific method shows its incomparable superiority over philosophical speculation. A sound set of postulates should lead to the deduction of theorems representing phenomena never previously investigated quite as logically as of theorems representing phenomena already known when the postulates were formulated. When a theorem representing novel phenomena receives direct observational confirmation there is no possi-

be futile. In science, on the other hand, the situation may be almost completely reversed; the conclusion (or theorem) may be known observationally at the outset, but the premises (or postulates) may at first be little more than conjectures and the logical process quite circular. For the methodology of resolving this circularity, see p. 8 ff.

[5] From the experimental point of view the process of developing systematic theory thus leads in two directions. On one hand it leads to the investigation of theorems derived from postulates of the system, and on the other to the direct investigation of postulates which appear to be required as assumptions for the deductive explanation of facts already known. Since phenomena of the latter type are fundamental in a strict sense, their investigation is of the highest significance. A background of systematic theory thus often directly suggests fundamental investigations which might be indefinitely delayed under the usual procedure of random, and even of systematic, exploration.

bility of circularity; as a consequence the probability that the postulates directly involved are sound is very definitely increased.[6] Thus the fact that Venus shows lunar phases could not have been known to Copernicus when he formulated his epoch-making hypothesis, because the telescope had not yet been invented. Accordingly their discovery by Galileo constituted strong positive evidence of the essential soundness of the Copernican hypothesis regarded as a postulate. This classical example of the observational but indirect confirmation of the soundness of postulates will serve as a fitting conclusion for our general consideration of theoretical methodology.

The Recognized Scientific Methodology Has Not Been Applied to the Behavior Controversy

We turn now to the question of whether the recognized scientific methodology is really applicable to a resolution of the controversy concerning the basic nature of adaptive behavior. At first glance the prospect is reassuring. It becomes quite clear, for example, what Weiss and Eddington should have done to substantiate their claims. They should have exhibited, as strict logical deductions from explicitly stated postulates, a series of theorems corresponding in detail to the concrete manifestations of the higher forms of human behavior. Then, and only then, they might proceed to the examination of the postulates of such system. To substantiate his position Weiss would have to show that these postulates concern essentially the behavior of electrons, protons, etc.; and Eddington to support his assertions would need

[6] A single unequivocal disagreement between a theorem and observed fact is sufficient to assure the incorrectness of at least one of the postulates involved. But even if the postulates of a system generate a very long series of theorems which are subsequently confirmed without exception, each new confirmation merely adds to the *probability* of the truth of such postulates as are incapable of direct observational test. Apparently this indirect evidence never reaches the crisp certainty of a deductive conclusion in which the postulates are directly established, except in the highly improbable situation where all the possible deductions involving a given postulate have been tested with positive results. According to the theory of chance, the larger the sample from this possible total which has been tried and found without exception to be positive, the greater the probability that a new deduction based on the same set of postulates will be confirmed when tested.

to show that the postulates of a successful system are primarily phenomena of consciousness.[7] The formal application of the methodology is thus quite clear and specific.

But here we meet an amazing paradox. In spite of the calm assurance of Weiss as to the truth of his statement that purposive behavior is at bottom physical, we find that he neither presents nor cites such a system. Indeed, he seems to be quite oblivious of such a necessity. Turning to Eddington, we find exactly the same paradoxical situation. Notwithstanding his positive, even emphatic, implications that moral behavior must be conscious or psychic in its ultimate nature, we find him neither presenting nor citing a theoretical system of any kind, much less one derived from psychic or conscious postulates. This paradox is particularly astonishing in the case of Eddington because he has been active in the field of physical theory and should, therefore, be sophisticated regarding the essential methodology involved in scientific theory in general. Surely the same logic which demands strict deduction from explicitly stated postulates in physical theory demands it for the theory of adaptive and moral behavior. And surely if we demand it of a mechanistic theory of the more recondite forms of human behavior, as Eddington seems emphatically to do, there is no hocus-pocus whereby a psychic view of such behavior may be maintained without the same substantial foundation.

A Demonstration of the Application of Theoretical Methodology to Adaptive Behavior

But if neither Weiss nor Eddington, nor any other writer in this field, has been able to bring forward the indispensable systematic theory as a prerequisite of the logical right to express a valid conclusion concerning the ultimate nature of higher adaptive behavior, may this not mean that the attainment of such a system is impossible, and that, consequently,

[7] It is here assumed as highly probable that if the two approaches are strictly in conflict, only one would be successful. In the course of the development of scientific theoretical systems, however, it is to be expected that during the early stages several different systems may present appreciable evidences of success. See The conflicting psychologies of learning—A way out, Psychol. Rev., 42, 1936; especially pp. 514–515.

the problem still remains in the realm of philosophical speculation? There is reason to believe that this is not the case. The ground for optimism lies in part in the small theoretical system which is presented below (p. 15).

By way of introduction to the system we may begin with the consideration of Theorem I (p. 17). In brief, this theorem purports to show that Pavlov's conditioned reactions and the stimulus-response 'bonds' resulting from Thorndike's so-called 'law of effect' are in reality special cases of the operation of a single set of principles. The major principle involved is given in Postulate 2. Briefly, this postulate states the assumption of the present system concerning the conditions under which stimuli and reactions become associated. The difference in the two types of reaction thus turns out to depend merely upon the accidental factor of the temporal relationships of the stimuli to the reactions in the learning situation, coupled with the implication that R_G, which in part serves to mark a reinforcing state of affairs, is also susceptible of being associated with a new stimulus.[8] The automatic, stimulus-response approach thus exemplified is characteristic of the remainder of the system.

A consideration of Theorem II will serve still further as an orientation to the system before us. We find this theorem stating that both *correct* and *incorrect* reaction tendencies may be set up by the conditioning or associative process just referred to. Our chief interest in this theorem, as an introduction to the system, concerns the question of whether the terms 'correct' and 'incorrect' can have any meaning when they refer to reaction tendencies which are the result of a purely automatic process of association such as that presented by Postulate 2. It is believed that they have a very definite meaning. Definitions 7 and 8 state in effect that correctness or incorrectness is determined by whether the reaction tendency under given conditions is, or is not, subject to experimental extinction. Such purely objective or behavioral

[8] In effect this deduction purports to show that the Pavlovian conditioned reflex is a special case under Thorndike's 'law of effect,' though Thorndike might not recognize his favorite principle as formulated in Postulate 2. For a fuller but less formal discussion of this point see *Psychol. Bull.*, 1935, 32, 817–822.

definitions of numerous terms commonly thought of as applying exclusively to experience, as distinguished from action, are characteristic of the entire system.

With this general orientation we may proceed to the theorems more specifically concerned with adaptive behavior. The proof of the first of these, Theorem III, shows that under certain circumstances organisms will repeatedly and successively make the same incorrect reaction. At first sight this may seem like a most commonplace outcome. However, when considered in the light of the definition of correctness given above it is evident that this theorem differs radically from what might be deduced concerning the behavior of a raindrop or a pebble moving in a gravitational field.[9]

Theorem IV states that after making one or more incorrect reactions an organism will spontaneously vary the response even though the environmental situation remains unchanged. This theorem is noteworthy because it represents the classical case of a form of spontaneity widely assumed, as far back as the Middle Ages, to be inconceivable without presupposing consciousness.

Theorem V states that when an organism originally has both correct and incorrect excitatory tendencies evoked by a single stimulus situation, the correct tendency will at length be automatically selected in preference to stronger incorrect ones.[10] This theorem, also, has been widely regarded as impossible of derivation without the presupposition of consciousness. Otherwise (so it has been argued) how can the organism know which reaction to choose?

[9] It may be suggested that if water should fall into a hollow cavity on its way to the sea, it might at first oscillate back and forth vigorously and then gradually subside, each oscillation corresponding to an unsuccessful attempt and the gradual cessation, to experimental extinction. In all such cases the discussion as to whether the observed parallelism in behavior represents an essential similarity or a mere superficial analogy requires that both phenomena possess a thorough theoretical basis. *If the two phenomena are deducible from the same postulates and by identical processes of reasoning, they may be regarded as essentially the same, otherwise not.* But if one or both lacks a theoretical basis such a comparison cannot be made and decision can ordinarily not be reached. Much futile argument could be avoided if this principle were generally recognized.

[10] See Simple trial-and-error learning: A study in psychological theory, Psychol. Rev., 1930, 37, 241–256; especially pp. 243–250.

Theorem VI represents the deduction that in certain situations the organism will give up seeking, *i.e.*, cease making attempts, and thus fail to perform the correct reaction even when it possesses in its repertoire a perfectly correct excitatory tendency. The substance of this proof lies in the expectation that the extinction resulting from repeated false reactions will cause indirectly a critical weakening of a non-dominant but correct reaction tendency. This theorem is of unusual importance because it represents the deduction of a phenomenon not as yet subjected to experiment. As such it should have special significance as a test of the soundness of the postulates.

With Theorems VII and VIII we turn to the problem of anticipatory or preparatory reactions. The proof of Theorem VII derives, from the principles of the stimulus trace and conditioning (Postulates 1 and 2), the phenomenon of the antedating reaction. The substance of this theorem is that after acquisition, learned reactions tend to appear in advance of the point in the original sequence at which they occurred during the conditioning process.[11] Pursuing this line of reasoning, Theorem VIII shows that in the case of situations demanding flight, such antedating reactions become truly anticipatory or preparatory in the sense of being biologically adaptive to situations which are impending but not yet actual. Thus we arrive at behavioral foresight, a phenomenon evidently of very considerable survival significance in animal life and one frequently regarded as eminently psychic, and inconceivable without consciousness.[12]

Passing over Theorem IX, which lays some necessary groundwork, we come to Theorem X. Here we find a deduction of the existence of the fractional anticipatory goal reaction. Of far greater significance from our present point of view, the deduction purports to show that through the action of mere association the fractional anticipatory reaction tends automatically to bring about on later occasions the

[11] See A functional interpretation of the conditioned reflex, Psychol. Rev., 1929, 36, 498–511; especially pp. 507–508.

[12] See Knowledge and purpose as habit mechanisms, Psychol. Rev., 1930, 37, 511–525; especially pp. 514–516.

state of affairs which acted as its reinforcing agent when it was originally set up. For this and other reasons it is believed that the anticipatory goal reaction is the physical basis of expectation, of intent, of purpose, and of guiding ideas.[13]

Theorem XI represents a deduction of the phenomenon of behavioral disappointment[14] as manifested, for example, by Tinklepaugh's monkeys. When these animals had solved a problem with the expectation of one kind of food they would tend to refuse a different kind of food, otherwise acceptable, which had been surreptitiously substituted.[15]

Theorem XII purports to be the deduction of the principle that organisms will strive actively to attain situations or states of affairs which previously have proved to be reinforcing. The automaticity deduced in the proof of Theorem X has here reached a still higher level. This is the capacity to surmount obstacles. But with the ability to attain ends in spite of obstacles comes automatically a genuine freedom (Definition 18), of great biological value but in no way incompatible with determinism.[16]

Theorem XIII is also derived with the aid of the fractional anticipatory goal reaction. This theorem represents the phenomenon of the adaptive but automatic transfer of learned reactions to situations having, as regards *external* characteristics, nothing whatever in common with the situations in which the habits were originally acquired. This, once more, is a form of adaptive behavior of the greatest survival significance to the organism, and one supposed in certain quarters to be impossible of derivation from associative principles.

[13] See Goal attraction and directing ideas conceived as habit phenomena, Psychol. Rev., 1931, 38, 487–506.

[14] It is to be observed from a comparison of Definitions 9 and 16 that *Disappointment* necessarily presupposes a specific expectation or intent (r_G), whereas *Discouragement* does not.

[15] O. L. Tinklepaugh, An experimental study of representative factors in monkeys, *J. Comp. Psychol.*, 1928, 8, 197–236. See especially p. 224 ff.

[16] An additional element of interest in this theorem is the fact that the fundamental phenomenon of motivation seems to have been derived from the ordinary principle of association (Postulate 2). If this deduction should prove to be sound, it will have reduced the two basic categories of motivation and learning to one, the latter being primary.

This is believed to be a low but genuine form of insight and a fairly high order of the 'psychic.'

This concludes the list of formally derived theorems. They have been selected from a series of fifty or so which are concerned with the same subject. None of these theorems 'reaches' Eddington's 'rational man morally responsible.' They accordingly are not offered as a basis for deciding the ultimate nature of such behavior. They *are* offered as a concrete and relevant illustration of the first and most essential step in the methodology which must be followed by Eddington, or anyone else who would determine the basic nature of the higher forms of behavior. Incidentally they are offered as specific evidence that such problems, long regarded as the peculiar domain of philosophy, are now susceptible of attack by a strictly orthodox scientific methodology.

Adaptive Behavior—A Scientific Theoretical System in Miniature[17]

Definitions

1. A *reinforcing state of affairs* (Postulate 3) is one which acts to give to the stimulus-trace component (Postulate 1) of preceding or following temporal coincidences consisting of a stimulus trace and a reaction, the capacity to evoke the reaction in question (Postulate 2).

2. *Experimental extinction* is the weakening of a conditioned excitatory tendency resulting from frustration or the failure of reinforcement (Postulate 4).

3. *Frustration* is said to occur when the situation is such that the reaction customarily evoked by a stimulus complex cannot take place (Postulate 4).

4. *Seeking* is that behavior of organisms in trial-and-error situations which, upon frustration, is characterized by varied alternative acts all operative under the influence of a common drive (S_D).

5. An *attempt* is a segment of behavior the termination of which is marked by either reinforcement or extinction.

6. A *simple trial-and-error situation* is one which presents to an organism a stimulus complex which tends to give rise to multiple reaction tendencies which are mutually incompatible, one or more of them being susceptible to reinforcement and one or more of them not being so susceptible.

7. A *correct* or 'right' reaction is a behavior sequence which results in reinforcement.

8. An *incorrect* or 'wrong' reaction is a behavior sequence which results in experimental extinction.

---

[17] The author is greatly indebted to Dr. E. H. Rodnick and Mr. D. G. Ellson for detailed criticisms and suggestions during the original preparation of the system which follows. Thanks are also due Professor K. F. Muenzinger, Dr. R. T. Ross, and Dr. R. K. White for criticisms given since the presentation at Hanover.

9. *Discouragement* is the diminution in the power of one excitatory tendency to evoke its normal reaction, this diminution resulting from one or more unsuccessful attempts involving a second reaction.

10. A behavior sequence is said to be *directed* to the attainment of a particular state of affairs when there appears throughout the sequence a characteristic component (r_G) of the action (R_G) closely associated with the state of affairs in question and this component action (r_G) as a stimulus tends to evoke an action sequence leading to the total reaction (R_G) of which the component constitutes a part.

11. Striving is that behavior of organisms which, upon frustration, displays varied alternative action sequences, all *directed* by an intent (r_G) to the attainment of the same reinforcing state of affairs.

12. A *goal* is the reinforcing state of affairs towards the attainment of which a behavior sequence of an organism may be directed by its intent (r_G).

13. An organism is said to *anticipate* a state of affairs when there is active throughout the behavior sequence leading to the state of affairs a fractional component (r_G) of the action associated with the state of affairs in question.

14. *Success* is the culmination of striving which is characterized by the occurrence of the full reaction (R_G) of which the fractional anticipatory component (r_G) is a part.

15. *Failure* is the culmination of striving which is characterized by the lack of the enactment of the full reaction (R_G) of which the fractional component (r_G) is a part.

16. *Disappointment* is the diminution in the power of one reinforcing situation to evoke appropriate consummatory reaction, this diminution (Postulate 4) resulting from the failure of a second reaction sequence directed (by an intent, or r_G) to a different reinforcing situation from that to which the first was directed, both being based on the same drive (S_D).

17. A *habit-family hierarchy* consists of a number of habitual behavior sequences having in common the initial stimulus situation and the final reinforcing state of affairs.

18. *Individual freedom* of behavior, so far as it exists, consists in the absence of external restraint.

Postulates

1. The adequate stimulation of a sense organ initiates within the organism a neural reverberation which persists for some time after the stimulus has ceased to act, the absolute amount of the reverberation diminishing progressively to zero but at a progressively slower rate. (Stimulus trace.)

2. When a reaction and a given segment of a stimulus-trace (Postulate 1) repeatedly occur simultaneously and this coincidence occurs during the action of a drive (S_D) and temporally close to a reinforcing state of affairs (Definition 1), this and stronger segments of the stimulus trace tend progressively to acquire the capacity to evoke the reaction, the strength of the association thus acquired manifesting a negatively accelerated diminution with distance of the associates from the reinforcing state of affairs. (Positive association.)

3. A characteristic stimulus-reaction combination ($S_G \longrightarrow R_G$) always marks reinforcing states of affairs (Definition 1). The particular stimulus-response combination marking the reinforcing state of affairs in the case of specific drives is determined empirically, *i.e.*, by observation and experiment. (Mark of reinforcing state of affairs.)

4. When a stimulus evokes a conditioned (associative) reaction (Postulate 2) and this event does not occur within the range of the reinforcing state of affairs (Defi-

nition 1 and Postulate 3), or when an excitatory tendency in a behavior sequence encounters a situation which makes the execution of the act impossible (Definition 3), the excitatory tendency in question undergoes a diminution in strength with a limit below the reaction threshold (Definition 2), this diminution extending in considerable part to other excitatory tendencies which may be operative at the same time or for some time thereafter. (Negative association or experimental extinction.)

5. The strength of any given increment of either positive or negative association (Postulates 2 and 3) diminishes with the passage of time, and the portion remaining shows a progressively greater resistance to disintegration with the increase in time since its acquisition, a certain proportion of each increment being permanent. (Negative retention or forgetting.)

6. Each reaction of an organism gives rise to a more or less characteristic internal stimulus. (Internal stimulation.)

Key to Diagrams

S = an adequate stimulus together with the resulting trace (Postulate 1).
S_D = the stimulus associated with a drive, such as hunger.
S_G = the stimulus associated with the goal or reinforcing state of affairs.
s = an internal stimulus resulting from a reaction.
R = a reaction.
R_G = the reaction associated with the goal or reinforcing state of affairs.
r_G = a fractional component of the goal reaction.
— — → = excitatory tendency from stimulus to reaction.

〰→ = causal connection of a non-stimulus-reaction nature.

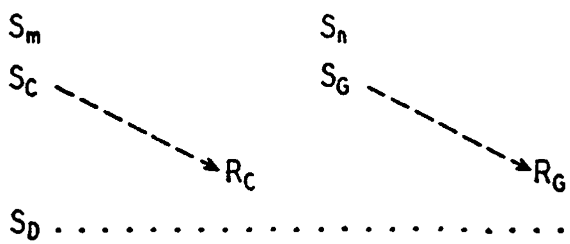
. . . = a continuation or persistence of a process, as of a drive (S_D).
Distance from left to right represents the passage of time.

Theorems

I

The Pavlovian conditioned reaction and the Thorndikian associative reaction are special cases of the operation of the same principles of learning.

1. Suppose that in the neighborhood of a sensitive organism stimuli S_C and S_G occur in close succession, that these stimuli in conjunction with the drive (S_D) evoke reactions R_C and R_G respectively, that S_m coincides in time with S_C while S_n coincides in time with S_G, and that (Postulate 1) the stimulus trace of S_m extends to R_C, and the stimulus trace of S_n extends to R_G.



2. From (1) and Postulate 1, it follows that one phase of the stimulus trace of S_m will coincide with R_C and one phase of the stimulus trace of S_n will coincide with R_G.

3. Now, by Postulate 3, the combination $S_G \dashrightarrow R_G$ marks a reinforcing state of affairs.

4. From (1), (2), (3), and Postulate 2 it follows, among other things, that the trace of S_n will become conditioned to R_G, and the trace of S_m will be conditioned to R_C, yielding the following excitatory tendencies:

$$S_m \dashrightarrow R_C$$
$$S_n \dashrightarrow R_G$$

5. But by (3) and (4) the reaction of the newly acquired excitatory tendency $S_n \dashrightarrow R_G$ is that intimately associated with the reinforcing state of affairs, which identifies it as a conditioned reaction of the Pavlovian type.

6. On the other hand, by (3) and (4) the reaction of the excitatory tendency $S_m \dashrightarrow R_C$ is a reaction distinct from that of the reinforcing state of affairs, which identifies it as an associative reaction of the Thorndikian type.

7. By (5) and (6) both the Pavlovian and the Thorndikian types of reaction have been derived from (1), (2), (3), and (4) jointly, and these in turn from the same principles of learning (Postulates 1, 2, and 3).

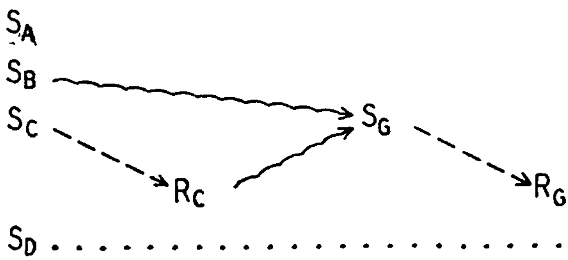
8. From (7) the theorem follows.

Q. E. D.

II

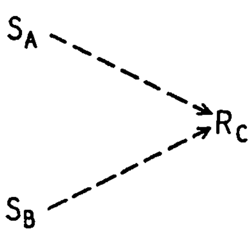
Both correct (right) and incorrect (wrong) reactions may be set up by the conditioning (associative) process.

1. Let it be supposed that an organism capable of acquiring associative reactions (Postulate 2) is, a number of times, stimulated simultaneously by S_A, S_B, S_C, and S_D; that S_C evokes reaction R_C; that the stimulus trace (Postulate 1) of S_A and S_B extend as far as R_C; that the object represented by S_B, in conjunction with act R_C, produces (causes) in the external world the event yielding the stimulus S_G; and finally that S_G evokes R_G.



2. By Postulate 3, $S_G \dashrightarrow R_G$ marks a reinforcing state of affairs.

3. From (1), (2), and Postulates 1 and 2, it follows that among other associative tendencies the following must be set up:



4. Now suppose that at a later time, S_B *alone* should evoke R_C. It follows from (1) that S_B, in conjunction with R_C, will cause to occur the event in the external world which will yield the stimulus S_G which, in turn, will evoke R_G.

5. But, once more, by Postulate 3, $S_G \dashrightarrow R_G$ marks a reinforcing state of affairs from which it follows that under the special new conditions of (4) the reaction tendency $S_B \dashrightarrow R_C$ will still be reinforced.

6. From (5) and Definition 7 it follows that $S_B \dashrightarrow R_C$ must be a correct or 'right' reaction.

7. Let us suppose, on the other hand, that S_A *alone* should evoke R_C. It follows from (1) that the external event giving rise to S_G will not occur (S_B being absent), and the excitatory tendency $S_A \dashrightarrow R_C$ will not be reinforced and, by Postulate 4, will suffer experimental extinction.

8. From (7) and Definition 8 it follows that $S_A \dashrightarrow R_C$ will be an incorrect or 'wrong' reaction.

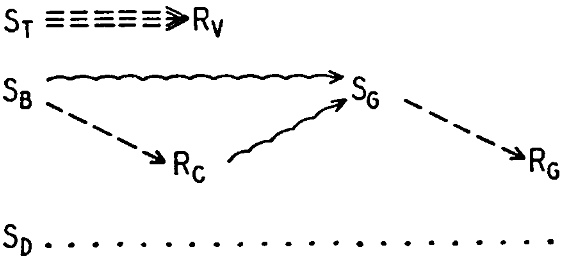
9. From (6) and (8) the theorem follows.

Q. E. D.

III

Simple trial-and-error situations may arise in which the organism will make repeated incorrect reactions.

1. Let it be supposed that we have the simultaneous stimulus situation $S_TS_BS_D$ with the component S_B (step 3, Theorem II) evoking R_C; that S_B and R_C when operating jointly cause S_G, S_G evoking R_G, whereas S_T evokes R_V with an excitatory tendency exceeding that of S_B to R_C by an amount greater than the weakening effect (Postulate 4) of several unreinforced attempts (Definition 5); that R_V is not followed by its usual reinforcing sequence ($S'_G \dashrightarrow R'_G$); and that the external stimulus situation after each attempt becomes exactly the same as before.



2. From (1) it follows that reaction R_V will take place at once after the organism encounters the compound stimulus $S_TS_BS_D$.

3. Now, by (1) the situation is such that R_V cannot be followed by its accustomed reinforcing sequence, so that this behavior sequence must be interrupted.

4. From (3) and Postulate 4, the excitatory tendency from S_T to R_V will be weakened by experimental extinction.

5. By (4) and Definition 8, R_V is an incorrect reaction.

6. By (1) and (2) the stimulus situation after the first R_V reaction must be the same as at the beginning, and the excitatory tendency to R_V must still be considerably in excess of that to R_C, from which it follows that R_V will occur a second time, and so on.

7. But, by (2), we have a first reaction (R_V), which, by (5) is incorrect and by (6) we have a repetition of this incorrect reaction, from which the theorem follows.

Q. E. D.

IV

Organisms in simple trial-and-error situations may manifest spontaneous variability of reaction, the objective situation remaining constant.

1. Suppose the situation in (1) of Theorem III with the additional assumption that excitatory tendency $S_B \dashrightarrow R_C$ shall be strong enough to resist all generalized inhibitory effects (Postulate 4) sufficiently to escape becoming subliminal.

2. By (1) and Theorem III, it follows that reaction R_V will take place repeatedly.

3. By (1) and (2), reaction $S_T \dashrightarrow R_V$ will not be followed by reinforcement (neither $S'_G \dashrightarrow R'_G$ nor $S_G \dashrightarrow R_G$), which failure (Postulate 4) will progressively weaken the tendency to R_V.

4. From (1) and (3) it follows that the reaction tendency to R_V must finally become weaker than that to R_C, at which point the stimulus complex $S_TS_BS_D$ will evoke reaction R_C.

5. But the shift from reaction R_V (2) to R_C (4) constitutes a variability of reaction.

6. Meanwhile, by (1) the objective situation has not changed.

7. From (5) and (6) the theorem follows.

Q. E. D.

V

Organisms in simple trial-and-error situations beginning with erroneous reactions may, after a sufficiently large number of attempts, come to give an indefinitely long series of successive correct reactions.

1. Let us assume the situation in step (1) of the deduction of Theorem IV.

2. By (1) and steps (2), (3), and (4) of Theorem IV, reaction tendency R_V will be progressively weakened by extinction until it is below the level of R_C, when the latter will take place.

3. Moreover, by (1), R_C in conjunction with S_B causes S_G; and S_G evokes R_G which, by Postulate 3, marks a reinforcing state of affairs.

4. It follows from (2), (3), and Postulate 2 that the excitatory tendency $S_B \dashrightarrow R_C$ will be reinforced, and therefore strengthened.

5. But a certain amount of time must elapse while reaction R_C is taking place; by Postulate 5, this time must permit a certain amount of spontaneous recovery from experimental extinction on the part of R_V.

6. Now, the rate of the spontaneous recovery of R_V (5) may be either (A) more rapid than the gain in strength of R_C through the latter's reinforcement, or (B) it may be less rapid, or (C) the two processes may take place at the same rate. If it is less rapid, or if the two processes take place at the same rate, R_C will maintain its dominance, thus giving an indefinitely long series of correct reactions (Definition 7); from which the theorem follows.

7. But suppose, on the other hand, that the rate of the spontaneous recovery of R_V from its experimental extinction is faster than the gain in strength of R_C through its reinforcement (6). It follows that on this alternative R_V must again become dominant.

8. From (7) it follows by reasoning analogous to that in (2) that R_V will occur repeatedly until depressed by further experimental extinction below the strength of R_C when the latter will again occur, to be further reinforced, and so on.

9. Now it follows from (4) and (8) together with Postulate 5, that after each complete cycle of reversal of R_V and R_C, the former will retain a certain amount of its weakening which will not yield to spontaneous recovery and the latter will retain a certain amount of the strengthening which will not yield to forgetting.

10. It follows from (9) that if the cyclical alternation were to go on indefinitely, the tendency to R_V must be weakened to zero and that to R_C must be strengthened to its maximum.

11. It is evident from (10) that at some point in the progressive shift in the basic strengths of R_V and R_C the two movements must cross, at which point R_C will be permanently dominant over R_V irrespective of spontaneous recovery or forgetting, and there will then follow an indefinitely long series of successive correct reactions.

12. From (6) and (11) the theorem follows.

Q. E. D.

VI

In simple trial-and-error learning situations, failure of final correct reaction will, under certain conditions, result from discontinued effort.

1. Suppose the situation in (1) of Theorem III except that the excitatory tendency $S_B \dashrightarrow R_C$ is at the outset only a little above the reaction threshold.

2. From (1) and Theorem III, false reaction R_V will be made repeatedly.

3. By (1) and (2), reaction tendency $S_T \dashrightarrow R_V$ will not be followed by reinforcement, which failure (by Postulate 4) will, if not interrupted, gradually weaken $S_T \dashrightarrow R_V$ to zero.

4. By (3) and Postulate 4, the weakening of $S_T \dashrightarrow R_V$ will extend in considerable part to $S_B \dashrightarrow R_C$.

5. Now, by (1) the super-threshold margin of strength of $S_B \dashrightarrow R_C$ may be smaller than any assigned finite value, from which it follows that it may be smaller than the depressing effects (4) arising from the extinction of $S_T \dashrightarrow R_V$.

6. It follows from (5) that before $S_T \dashrightarrow R_V$ may be extinguished beneath the level of $S_B \dashrightarrow R_C$ the latter will also have been depressed below the reaction threshold so that when $S_T \dashrightarrow R_V$ reaches zero and ceases action, the potentially correct reaction tendency, $S_B \dashrightarrow R_C$, will also be unable to function even though without any competition whatever.

7. But the depression of both the tendency to R_V and R_C as shown in (6) will bring about a cessation of attempts (Definition 5), the latter of which (1) would have been a correct reaction (Definition 7).

8. From (3), (4), and (7) the theorem follows.

Q. E. D.

Corollary 1.

Organisms capable of acquiring competing excitatory tendencies will manifest discouragement.

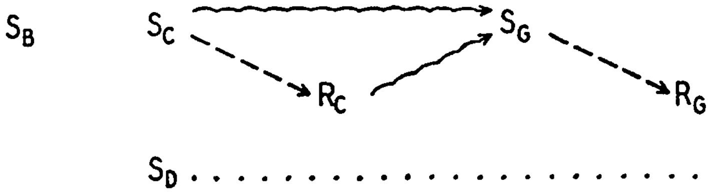
This follows directly from Theorem VI and Definition 9.

VII

Reactions conditioned to a late segment of a stimulus trace will subsequently occur as antedating reactions.

1. Suppose that stimulus S_B precedes stimulus S_C by several times the latency of conditioned reactions; that S_C evokes reaction R_C; that the stimulus trace of S_B extends as far as R_C; that the physical event responsible for S_C, jointly with reaction

R_C, causes S_G; that S_G evokes R_G; and that S_D begins at S_C and persists throughout the remainder of the process.



2. From (1) and Postulate 1, a segment of the stimulus trace initiated by S_B will coincide with R_C.

3. By (1) and Postulate 3, $S_G \dashrightarrow R_G$ marks a reinforcing state of affairs and follows the coincidence of R_C with the trace of S_B.

4. By (2), (3), and Postulate 2, reaction R_C will become conditioned to a late coinciding segment of the trace of stimulus S_B, *i.e.*, that portion which coincides temporally with R_C.

5. Now, by Postulate 1, stimulus trace S_B at the point of the onset of the stimulus is substantially the same as at the segment conditioned to R_C, except that it is stronger.

6. From (5) and Postulate 2 it follows that once R_C has been conditioned to a late segment of the trace of stimulus S_B with a supraliminal strength, the reaction will be evoked by any portion of the same trace which is as strong as, or stronger than, the segment conditioned.

7. But since, by (1), the initial portion of the stimulus trace of S_B will occur several times the latency of such a reaction in advance of the original point of the occurrence of R_C, it follows from (5) and (6) that after conditioning, R_C will be evoked in advance of the point of its original occurrence.

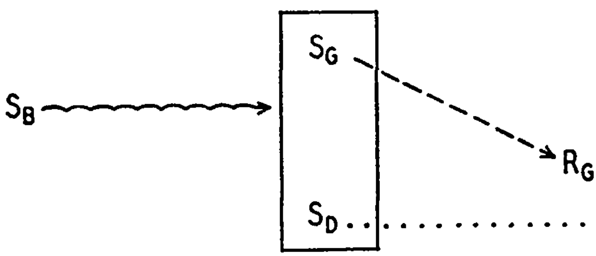
8. From (7) the theorem follows.

Q. E. D.

VIII

Organisms capable of acquiring trace conditioned reactions will be able to execute successful defense reactions.

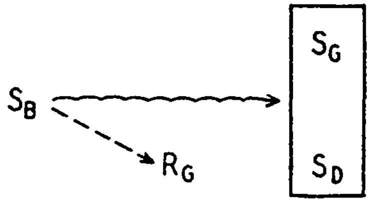
1. Let it be supposed that an organism capable of acquiring trace conditioned reactions is stimulated by S_B, that the external world event responsible for S_B initiates a causal sequence several times the length of a conditioned reaction latency, which sequence terminates in S_G and S_D, the two latter jointly constituting an injury and evoking R_G, a flight reaction, which terminates their impact on the organism; and that the stimulus trace of S_B reaches well beyond the point at which R_G occurs.



2. From (1) and Postulates 1 and 2 it follows that R_G will be conditioned to the trace of S_B.

3. From (1), (2), and Theorem VII it follows that if S_B occurs on a later occasion, reaction R_G will occur in advance of situation S_GS_D, which, if it impinges on the organism, will be injurious.

4. But, by (1), R_G is a flight reaction. It follows from (3) that the organism will not be present when the situation otherwise giving rise to S_GS_D occurs and so will escape the injury, thus:



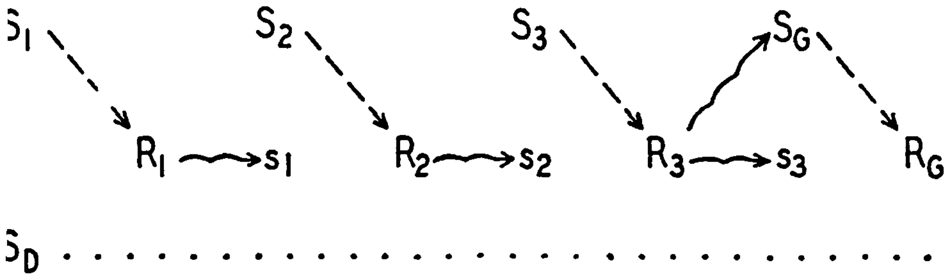
5. From (1) and (4) the theorem follows.

Q. E. D.

IX

In stable behavior sequences terminating in reinforcement, each reaction, in general, becomes conditioned (A) to the proprioceptive stimulus arising from the action immediately preceding it, and (B) to the drive stimulus (S_D), each with an intensity diminishing according to a negatively accelerated rate with distance from the reinforcing state of affairs.

1. Let it be supposed that there impinges on an organism a uniform sequence of external stimuli S_1, S_2, S_3, etc.; that these stimuli evoke in the organism reactions R_1, R_2, R_3, etc.; that these reactions produce (Postulate 6) proprioceptive stimuli s_1, s_2, s_3, etc.; that R_3 by an external causal sequence produces a state of affairs which includes S_G; that S_G evokes R_G; that the combination $S_G \longrightarrow R_G$ marks (Postulate 3) a reinforcing state of affairs; and that throughout the sequence there occurs the persisting drive stimulus S_D.



2. By Postulates 1 and 2 the situation supposed in (1) will give rise to an association between each proprioceptive stimulus and the reaction immediately following thus:

$$s_1 \dashrightarrow R_2$$
$$s_2 \dashrightarrow R_3$$
$$s_3 \dashrightarrow R_G$$

3. Also, since by (1) S_D occurs at every point throughout the series, it follows from (1) and Postulate 2 that S_D will be conditioned to every reaction in the series, thus:

$$\begin{tikzcd}[row sep=small]
 & R_1 \\
 & R_2 \\
S_D \arrow[uur, dashed] \arrow[ur, dashed] \arrow[dr, dashed] \arrow[ddr, dashed] & \\
 & R_3 \\
 & R_G
\end{tikzcd}$$

4. By (2) and Postulate 2,

$$s_3 \dashrightarrow R_G > s_2 \dashrightarrow R_3 > s_1 \dashrightarrow R_2$$

and

$$(s_3 \dashrightarrow R_G) - (s_2 \dashrightarrow R_3) > (s_2 \dashrightarrow R_3) - (s_1 \dashrightarrow R_2)$$

5. Also by (3) and Postulate 2,

$$S_D \dashrightarrow R_G > S_D \dashrightarrow R_3 > S_D \dashrightarrow R_2 > S_D \dashrightarrow R_1$$

and

$$(S_D \dashrightarrow R_G) - (S_D \dashrightarrow R_3) > (S_D \dashrightarrow R_3) - (S_D \dashrightarrow R_2) > (S_D \dashrightarrow R_2) - (S_D \dashrightarrow R_1)$$

6. But the expressions in (4) and (5) represent negatively accelerated excitatory gradients diminishing with distance from the reinforcing state of affairs.

7. From (2), (3), and (6) the theorem follows.

Q. E. D.

X

A fractional anticipatory goal reaction as a stimulus will tend to bring about the reinforcing state of affairs with which the total goal reaction, of which it is a constituent part, is associated.

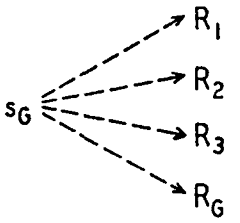
1. Suppose the situation in (1) of Theorem IX with the additional assumption that the goal reaction (R_G) is composed of two components, a major one which cannot take place without the aid of the object represented by S_G and which is incompatible with the several acts of the sequence preceding it, and a minor one (r_G) which is not mechanically dependent on S_G and which may take place simultaneously with the antecedent reactions of the series.

2. Now, by Theorem IX, S_D is conditioned to R_G and, since by (1) r_G is a constituent part of R_G, S_D is also conditioned to r_G.

3. Since, by (1), S_D occurs throughout the series, it follows that it will evoke r_G at all points in the behavior sequence R_1, R_2, R_3, etc.

4. From (3) and Postulates 1 and 6 it follows that the trace of the internal stimulus produced by r_G, *i.e.*, s_G, will tend to occur in conjunction with all the reactions of the sequence R_1, R_2, R_3, etc.

5. Now, each time the situation represented in (4) occurs it is followed (1) by the reinforcing state of affairs marked by $S_G \dashrightarrow R_G$, from which it follows by Postulate 2 that s_G will ultimately become associated with all of the reactions of the sequence, thus:



very much as in the case of S_D (Theorem IX).

6. From (5) it follows that s_G will tend, on subsequent occasions, to bring about reactions R_1, R_2, R_3. By (1), R_3 causes S_G, and S_G evokes R_G.

7. But by (1), r_G is a constituent part of R_G which, with its S_G, marks (*i.e.*, is associated with) the reinforcing state of affairs.

8. But if (6 and 7) r_G, through the action of s_G, brings about the inevitable mark of its reinforcing state of affairs ($S_G \dashrightarrow R_G$), it must at the same time bring about the reinforcing state of affairs itself.

9. From (7) and (8) the theorem follows.

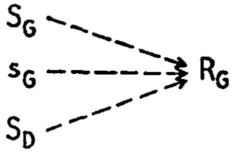
Q. E. D.

XI

Organisms capable of acquiring functionally potent anticipatory reactions intimately associated with the reinforcing state of affairs, will manifest a weakened tendency to the consummatory reaction if, at the completion of the action sequence, the state of affairs then presented does not permit the occurrence of the complete reaction of which the anticipatory reaction is a constituent part.

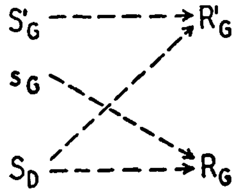
1. Suppose that an organism which has been in a situation such as (1) in Theorem IX later finds itself in the same situation with the exception that the terminal conditions, instead of permitting reaction $S_G \dashrightarrow R_G$, permit a different reaction, $S'_G \dashrightarrow R'_G$, which is appropriate to the same drive (S_D) and is in the repertoire of the organism in question but has a strength only slightly above the reaction threshold.

2. By step (1) of the proof of Theorem IX, together with Theorem IX itself, the customary stimulus complex giving rise to the terminal reaction must be:



3. Now, by (1) and Theorem IX, the s_G of (2) represents r_G, and r_G (Definition 13) is both a reaction anticipatory of, and a fractional component of, R_G.

4. On the other hand, by (1), (2), and Theorem IX, the excitatory tendencies under the changed conditions of the present theorem will be:



5. From (1) and (4) it follows that the excitatory tendencies leading to R_G must be frustrated (Definition 3) which (by Postulate 4) will set up experimental extinction at the point in question.

6. From (5) and Postulate 4 it follows that as a result of the extinction of the tendencies to R_G there will occur simultaneously a weakening of the tendency to reaction R'_G.

7. But by (1) the tendencies to R'_G may be as small as desired and therefore smaller than the generalized extinction of (6), from which it follows that under these circumstances the excitatory tendencies to R'_G will pass below the reaction threshold.

8. From (3) and (7) the theorem follows.

Q. E. D.

Corollary 1.

Organisms will display disappointment.

This follows directly from Theorem XI and Definition 16.

XII

Organisms capable of acquiring anticipatory goal reactions will strive to bring about situations which are reinforcing.

1. Let it be assumed that an organism has acquired a habit-family hierarchy (Definition 17) of two distinct action sequences of the type described in (1) of Theorem X, both originating in the external stimulus situation S_1, terminating in the reinforcing situation $S_G \longrightarrow R_G$ and associated with the drive S_D; that the initial acts of one of the sequences are R_I, R_{II}, etc., and those of the other are R_1, R_2, etc.; that the excitatory tendency initiating the sequence beginning with R_1 is dominant over that beginning with R_I, but that the tendency to R_I is far enough above the reaction threshold to survive the weakening effect which would result (Postulate 4) from the frustration of the tendency to R_1.

2. Now suppose that an obstacle is interposed which effectually prevents the completion of R_1 and the remainder of that sequence (1). It follows from Postulate 4 that this excitatory tendency will suffer extinction, with no limit above zero.

3. From (1) and (2) it follows that the sequence beginning with R_I and terminating with R_G will be executed after the frustration of the excitatory tendency leading to R_1.

4. Now, from (1) it follows by reasoning strictly analogous to steps (2), (3), (4), (5), and (6) of the deduction of Theorem X, that s_G will acquire during the acquisition of the habit family the tendency to evoke (*A*) reaction sequence R_1 and all those acts following it in the sequence leading to R_G, and (*B*) reaction sequence R_I together with all those leading from it to R_G.

5. From (2), (3), (4), and Definition 10 it follows that under these circumstances the introduction of a barrier will cause the organism to shift from one behavior sequence *directed* to a reinforcing state of affairs to another *directed to the same* reinforcing state of affairs.

6. But by (5) and Definition 11, when the interposition of an obstacle leads an organism to choose an alternative action sequence *directed* to the same reinforcing state of affairs as that interrupted by the obstacle, the behavior in question is striving.

7. From (6) the theorem follows.

Q. E. D.

Corollary 1.

Organisms will strive for goals.

This follows directly from Theorem XII and Definitions 11 and 12.

XIII

When an organism has attained a reinforcing state of affairs in a situation which, objectively considered, is totally novel, but by means of a member of a previously established habit-family hierarchy, there may follow without specific practice a tendency to a transfer to the new situation of the behavior tendencies represented by one or another of the remaining members of the habit-family hierarchy in question.

1. Let it be assumed that an organism has acquired a habit-family hierarchy (Definition 17) of two distinct action sequences of the type described in (1) of Theorem X, both originating in the external stimulus situation S_1, terminating in the reinforcing situation $S_G \dashrightarrow R_G$ and associated with the drive stimulus S_D; that the initial act of one of the sequences is R_I and that of the other is R_1.

2. From (1) it follows by reasoning strictly analogous to steps (2), (3), (4), (5), and (6) of the deduction of Theorem X, that s_G will acquire during the acquisition of the habit family the tendency to evoke (*A*) reaction sequence R_1 and all those acts following it in the sequence leading to R_G, and (*B*) reaction sequence R_I together with all those leading from it to R_G.

3. Now, suppose that this same organism in a novel external situation S'_1 and acting under the same drive stimulus S_D reaches, a few times, by the process of trial and error the reinforcing state of affairs marked by $S_G \dashrightarrow R_G$, by an action sequence the same as that beginning with R_1 of one of the members of the habit-family hierarchy of (1).

4. From (3) it follows by reasoning similar to steps (2) and (3) of the deduction leading to Theorem X, that r_G will be present throughout the behavior sequence beginning with S'_1.

5. It follows from (4) that there will be a coincidence of r_G and the stimulus trace of S'_1.

6. Since by (3) the coincidence of the stimulus trace of S'_1 and r_G (5) is followed by $S_G \dashrightarrow R_G$, it follows by Postulates 1 and 2 that there will be set up the excitatory tendency $S'_1 \dashrightarrow r_G$.

7. From (6), (2), and Postulate 6 it follows that S'_1 will tend to initiate the behavior sequence (omitting internal stimuli after R_1):

$$S'_1 \dashrightarrow r_G \rightsquigarrow s_G \dashrightarrow R_1 \dashrightarrow R_2 \dashrightarrow R_3 \rightsquigarrow S_G \dashrightarrow R_G$$

and also

$$S'_1 \dashrightarrow r_G \rightsquigarrow s_G \dashrightarrow R_{\mathrm{I}} \dashrightarrow R_{\mathrm{II}} \dashrightarrow R_{\mathrm{III}} \rightsquigarrow S_G \dashrightarrow R_G$$

or, combining the two sequences,

$$\begin{tikzcd}[column sep=small]
 & & & R_1 \arrow[r, dashed] & R_2 \arrow[r, dashed] & R_3 \arrow[dr, squiggly] & & \\
S'_1 \arrow[r, dashed] & r_G \arrow[r, squiggly] & s_G \arrow[ur, dashed] \arrow[dr, dashed] & & & & S_G \arrow[r, dashed] & R_G \\
 & & & R_{\mathrm{I}} \arrow[r, dashed] & R_{\mathrm{II}} \arrow[r, dashed] & R_{\mathrm{III}} \arrow[ur, squiggly] & &
\end{tikzcd}$$

8. Now, suppose that at this point an obstacle is interposed such that R_1 cannot take place, and that $s_G \dashrightarrow R_{\mathrm{I}}$ is far enough above the reaction threshold to resist the weakening effect of the frustration of the excitatory tendency to R_1. With the competition of R_1 thus removed from the excitatory tendency to R_{I}, s_G will initiate R_{I} (7); this will lead to R_{II}, this to R_{III}, etc., and finally to $S_G \dashrightarrow R_G$.

9. But the shift from the sequence beginning with R_1 to that beginning with R_{I} as in (8) is a transfer without specific practice from an old to a new situation because R_{I}, R_{II}, etc., have never taken place in the external stimulus situation beginning with S'_1.

10. From (9) the theorem follows.

Q. E. D.

The Nature of Adaptive Behavior as Indicated by the Postulates of the Present System

We come now to the second step in our exposition of the procedure which should have been carried out by Eddington and Weiss before they presumed to state the ultimate nature of the more complex forms of adaptive and moral behavior. In this step we turn, mainly for purposes of illustration, to the direct examination of the postulates which gave rise to the system, to see whether they are, in fact, physical or psychic. Let us pass them in review. Postulate 1 states that the physiological effects of a stimulus persist for a certain time even after the stimulus has ceased. Postulate 2 indicates the conditions under which stimuli and reactions become associated or conditioned. Postulate 3 gives the marks of reinforcing situations. Postulate 4 states the conditions under which associations are unlearned. Postulate 5 gives the conditions under which positive and negative learning are lost. Postulate 6 states the well-known fact of internal stimulation.

At first glance most persons would probably say that these postulates represent the behavior of what has always been regarded as physical. Moreover, the postulates appear to be phenomena of physical structures which most theoretical physicists believe will ultimately be derived, *i.e.*, deduced, by them from electrons, protons, deutrons, etc. According to this view the theoretical physicists will ultimately deduce as theorems from electrons, protons, etc., the six postulates which we have employed as the basis for the deduction of adaptive behavior. If this deduction were accomplished we should have an unbroken logical chain extending from the primitive electron all the way up to complex purposive behavior. Further developments may conceivably extend the system to include the highest rational and moral behavior. Such is the natural goal of science. This is the picture which a complete scientific monism would present. Unfortunately, theoretical physics is very far from this achievement, and judgment regarding its ultimate accomplishment must be indefinitely suspended. At most such a view, attractive as it is, can be regarded only as a working hypothesis.[18]

[18] There is conceivable, however, a kind of experimental shortcut to the determination of the ultimate nature of adaptive behavior. Suppose it were possible to construct from inorganic materials, such as the theoretical physicists have already succeeded in deriving from electrons and protons, a mechanism which would display exactly the principles of behavior presented in the six postulates just examined. On the assumption that the logic of the above deductions is sound, it follows inevitably that such a 'psychic' machine, if subjected to appropriate environmental influences, must manifest the complex adaptive phenomena presented by the theorems. And if, upon trial, this *a priori* expectation should be verified by the machine's behavior, it would be possible to say with assurance and a clear conscience that such adaptive behavior may be 'reached' by purely physical means. A beginning in the direction of such constructions has already been made. See R. G. Krueger and C. L. Hull, An electro-chemical parallel to the conditioned reflex, *J. Gen. Psychol.*, 1931, 5, 262–269; G. K. Bennett and L. B. Ward, Synthesis of conditioned reflex, *Amer. J. Psychol.*, 1933, 45, 339; D. G. Ellson, A mechanical synthesis of trial-and-error learning, *J. Gen. Psychol.*, 1935, 13, 212–218.

But What of Consciousness?

But what of consciousness, of awareness, of experience—those phenomena of which the philosophers and theologians have made so much and upon the priority of which they are so insistent? An inspection of the postulates of the miniature system of adaptive behavior presented above certainly shows no trace of any such phenomena. It is clear, therefore, that so far as that considerable array of complex behavior is concerned, consciousness or experience has no logical priority. In the field of scientific theory no other form of priority is of primary significance.

What, then, shall we say about consciousness? Is its existence denied? By no means. But to recognize the existence of a phenomenon is not the same thing as insisting upon its basic, *i.e.*, logical, priority. Instead of furnishing a means for the solution of problems, consciousness appears to be itself a problem needing solution. In the miniature theoretical system, no mention of consciousness or experience was made for the simple reason that no theorem has been found as yet whose deduction would be facilitated in any way by including such a postulate. Moreover, we have been quite unable to find any other scientific system of behavior which either has found consciousness a necessary pre-supposition or, having assumed it, has been able to deduce from it a system of adaptive behavior or moral action.[19] There is, however, no reason at all for not using consciousness or experience as a postulate in a scientific theoretical system if it clearly satisfies the deductive criteria already laid down.

[19] It is rather hoped and expected that this statement will be challenged. In the interest of the clarification of an important problem, it is desirable that the challenge be accompanied by a formal exhibition of the structure of the system supposed to manifest the critical characteristics. As illustrated above, a theoretical system is a considerable sequence of interlocking theorems, all derived from the same set of postulates. Too often what pass as systems in psychology are merely informal points of view containing occasional propositions which, even if logically derived, would be nothing more than isolated theorems. Some authors are prone to the illusion that such propositions could be deduced with rigor in a few moments if they cared to take the trouble. Others assert that the logic has all been worked by them 'in their heads,' but that they did not bother to write it out; the reader is expected to accept this on faith. Fortunately, in science it is not customary to base conclusions on faith.

If such a system should be worked out in a clear and unambiguous manner the incorporation of consciousness into the body of behavior theory should be automatic and immediate. The task of those who would have consciousness a central factor in adaptive behavior and in moral action is accordingly quite clear. They should apply themselves to the long and grinding labor of the logical derivation of a truly scientific system. Until such a system has been attained on a considerable scale, the advancement of science will be favored by their limiting their claims to statements of their hopes and wishes as such. Meanwhile, one cannot help recalling that for several centuries practically all psychological and philosophical theorists have set out precisely with the assumption of the priority of consciousness or experience. Considering the practically complete failure of all this effort to yield even a small scientific system of adaptive or moral behavior in which consciousness finds a position of logical priority as a postulate, one may, perhaps, be pardoned for entertaining a certain amount of pessimism regarding such an eventuality.

In view of the general lack of the kind of evidence which would be necessary to show the logical priority of consciousness, it may naturally be asked why there is such insistence upon its central significance. While there are many contributing factors, it can scarcely be doubted that an important element in the situation is found in the perseverative influences of medieval theology. During the Middle Ages, and for centuries thereafter, social or moral control was supposed to be effected largely through promises of rewards or punishments after death. Therefore something had to survive death to reap these rewards. Consciousness as a non-physical entity was considered incorruptible and thus immune to the disintegration of the flesh. Consequently it offered a logical possibility of something surviving physical death upon which scores might be evened among the shadows beyond the river Styx. But to be convincing, it was necessary for the thing rewarded or punished to be an essentially causal element in the determination of moral conduct or behavior. Thus it was imperative not only that consciousness be non-physical, but also that it be the basic factor in determining action. Such a view is incompatible with the belief that the more complex forms of human behavior could be derived without any reference whatever to consciousness. Tradition is strong, especially when fostered by powerful institutions. Accordingly, the frequent insistence on the logical priority of consciousness is not surprising, even when coming from persons who have no clear notion as to the origin of their feelings in the matter.

Thus it can hardly be doubted that psychology in its basic principles is to a considerable degree in the thrall of the Middle Ages, and that, in particular, our prevailing systematic outlook in the matter of consciousness is largely medieval. The situation depicted in a remarkable panel of the fresco by Orozco in the Dartmouth Library gives a powerful artistic representation of this. There, lifeless skeletons in academic garb assist solemnly at the gruesome travail of a reclining skeleton in the act of reproducing itself. What a picture of academic sterility! Fortunately the means of our salvation is clear and obvious. As ever, it lies in the application of scientific procedures. The methodology is old and tried; it goes back even to the time of Galileo. The present paper is, in reality, an exposition of the specific application of this technique in a systematic manner to the problems of complex adaptive behavior. Galileo practiced this methodology at the imminent risk of imprisonment, torture, and death. For us to apply the methodology, it is necessary only to throw off the shackles of a lifeless tradition.