3 HUMAN MEMORY AND THE CONCEPT OF REINFORCEMENT[1]

RICHARD C. ATKINSON and THOMAS D. WICKENS[2]
Stanford University

The purpose of this paper is to offer a theory about the role of reinforcement in human learning and to evaluate the theory against data from several different types of experiments. It should be emphasized that this analysis is restricted to human learning. Our discussion of reinforcement will be based on a more general theory of memory (Atkinson and Shiffrin, 1968a) that has been derived primarily from results of verbal-learning experiments. The remarks that we shall make about reinforcement have not been applied outside of this context, and accordingly we

[1]Preparation of this paper was supported by the National Aeronautics and Space Administration, Grant NGR-05-020-036, and by a fellowship from the National Institutes of Health to the second author. The authors would like to thank several of their co-workers who carried out experiments reported here; each will be acknowledged at the appropriate point in the paper. Thanks are also due to William Cook, Geoffrey Loftus, and John Schnorr for a critical reading of an earlier draft of this paper.

[2]Now at the University of California, Los Angeles.

are unwilling at this time to extrapolate the analysis to animal learning.

In his discussion of the law of effect, Thorndike (1931) proposed two alternative views regarding the nature of reinforcement. One view, which he favored, assumed that the action of a reinforcement produced a direct and automatic strengthening of stimulus-response associations. The other view, which Thorndike considered and rejected, postulated that reinforcement did not affect learning per se, but rather determined the choice of a response once the subject recalled the relevant events that had occurred on preceding trials of the experiment. These two alternative views have been maintained in the literature since that time, and much research has been done in an attempt to determine which is the true state of affairs (for an excellent review of this research see Postman, 1962). This distinction may be useful in a general way to categorize theories of reinforcement, but it is becoming increasingly clear that the set of theories qualifying in each category is so large and variegated that it is not possible to formulate experimental tests which meaningfully differentiate between them. With this reservation in mind, it still seems worth noting that we regard our discussion of reinforcement as most closely allied to the second of the two views. Thus our analysis is in general accord with the theorizing of Tolman (1932) and with the more recent analyses offered by Estes (1969) and by Buchwald (1969).

Our discussion of learning and memory is in terms of information-processing concepts (Broadbent, 1963; Simon and Newell, 1964). Accordingly, we view the processes involved in learning as an exchange and transfer of information between a number of memory storage units. The nature of these transfers and the properties of the storage units will be specified in some detail, but we offer no speculations about their inner structure or possible physiological representations. In our view, learning involves the transfer of information generated by sources both external and internal to the organism into some form of memory store that can hold it until it is needed later. Reinforcement is a modulation of this information flow. A reinforcing event, in this sense, serves two functions: first, to set in motion the processes that cause the transfer to take place, and second, to select what information is to be transferred. When the study of some item occurs in an experiment, information associated with it is coded and transferred to the subject's memory. In order to produce a response at a later point in time, this information must be retrieved by a process which involves a more or less active search of memory. Thus, the operations involved in a typical learning situation can be divided into two classes, one associated with storage and the other with retrieval of information from memory. In many experiments this distinction is reflected in the study and test phases of a trial. The distinction between storage and retrieval is fundamental to the system and is reflected in our analysis of reinforcement.

Reinforcement manipulations that affect the storage process are the ones most commonly studied. Indeed, typically when the term reinforcement is used, it refers to operations that cause information about events which have taken place (including, perhaps, the reinforcing event itself) to be stored. To understand how transfer is effected, it is necessary to realize that a reinforcing event plays two separate and distinct roles in determining the storage of information: an informational role and an attentional role.

The first concerns the knowledge that is provided by giving feedback to the subject about whether or not his response to a particular stimulus was correct. When a subject is told that his response was, for example, correct, this provides the information that he must store to assure correct performance on subsequent trials. The quality of this feedback can be varied in a number of ways, most obviously by varying the amount of information provided to the subject after an error. The use of a correction procedure, in which the subject is told the response that should have been made after an error, makes more information available than does a partial correction or a noncorrection procedure in which the correct response is not completely specified (Bower, 1962; Keller, Cole, Burke, and Estes, 1965; Millward, 1964). The quality of information provided by the feedback also can be manipulated by introducing a delay between the subject's response and this feedback. Under these conditions, some information about the situation may be lost or confused, so that the feedback information, when presented, is of less value.

The attentional component of reinforcement in the storage process is closely related to conventional ideas of reward. Reinforcement, in this sense, acts to direct the subject's attention to one aspect of the situation and not to others. Thus, when a reward is associated with certain items presented for learning and not with others, more study may be given to the rewarded items and consequently they may be learned more rapidly (Harley, 1965a). Indeed, we postulate that this is the principal role of incentives when presented at the time of study: to cause the subject to attend to certain items or aspects of the situation more intensely than to others.

The storage aspects of reinforcement have received a good deal of study. The same cannot be said about the role of reinforcement in the retrieval of information and the production of a response. Again, we believe that these effects can take at least two forms. On the one hand, when the payoff value associated with a particular item is presented at the time of study, it may become part of the information complex placed in memory and may even determine where in memory it is stored. If this is the case, storage for an item with a high payoff value, for example, will be different in some way from storage of an item with low payoff. Knowledge given at the time of test regarding the payoff value assigned to the item, therefore, can aid the subject by indicating where in memory to look and hence cause him to set up a more effective search. The other effect that reinforcement may have on retrieval is to dictate the effort and time the subject is willing to spend in searching memory. It often happens that the information necessary to produce a response may be available in memory, but for various reasons cannot be recovered without an extended search. Presumably, when items are presented for test which have been assigned high payoff values, the subject will engage in a more extensive search and hence will be more likely to retrieve the appropriate information. Unfortunately, these two effects are largely speculative and have not been carefully documented experimentally. We have, however, undertaken some preliminary studies, which will be described later, on reinforcement effects during retrieval.

The main body of this paper is divided into two sections. The first develops the theoretical system, and the second deals with applications of the theory to a number of experimental situations. The theoretical section begins with a fairly extensive discussion of the structure of human memory. Although this discussion will not explicitly consider the question of reinforcement, the nature of the reinforcing process is so much determined by how the subject uses his memory that it cannot be analyzed without first considering these more basic processes. As we have noted above, the action of reinforcement may be thought of, in part, as an attentional process. Accordingly, the second step in our analysis specifies more exactly the ways in which attention acts within the framework of the theory. This consideration brings us in turn to a discussion of reinforcement.

In the second section the theory is applied to a number of experiments involving the manipulation of reinforcement variables. The first of these demonstrates the workings of the memory system when items are given varying numbers of reinforcements under different presentation schedules. This example will also illustrate a number of the complexities that can plague an analysis of reinforcement: in particular, the ways in which the short- and long-term properties of memory can lead to apparently contradictory effects. The second application examines delay of reinforcement and illustrates how this variable can have many different effects depending on the precise conditions of learning. The role of feedback in learning will be examined in another way as part of a third experiment, using a concept-identification paradigm. One of the primary purposes of this discussion is to demonstrate that the actual responses made by a subject frequently fail to provide an adequate indicator of the reinforcing processes involved. The experiment will also show how superficially similar reinforcements can have markedly different effects, depending upon the strategy used by the subject. Finally, the last set of experiments considers the ways in which reward magnitude can lead to selective study of certain items and, in turn, affect both the storage and retrieval of information.

Before starting our discussion, a warning should be added. We view reinforcement as a complex process and one which is derived from other, more fundamental aspects of the learning situation. Because of this fact, the effects of reinforcement are often quite varied, both in their appearance and in the manner by which they are produced. Our discussion, therefore, may well prove unsatisfactory to someone who is looking for a single, unified law to explain all reinforcement phenomena. Such a law, we feel, does not exist.

Theoretical System

The memory system. Although the theory on which our discussion of reinforcement will be based has been described in other papers (Atkinson and Shiffrin, 1965, 1968a,b; Shiffrin and Atkinson, 1969), a brief review will provide a starting point for the work to be presented. This discussion will not present the theory in its full detail. In particular, no attempt will be made to consider all of the possible variants of the memory system, nor will explicit mathematical predictions of the theory be derived. For these matters, and for a description of the evidence which supports this formulation, the reader is referred to the previously cited theoretical papers and to reports of related experimental work (Atkinson, 1969; Atkinson, Brelsford, and Shiffrin, 1967; Brelsford and Atkinson, 1968; Brelsford, Shiffrin, and Atkinson, 1968; Freund, Loftus, and Atkinson, 1969; Phillips, Shiffrin, and Atkinson, 1967; Rundus, 1970; Rundus and Atkinson, 1970; Shiffrin, 1968; Thomson, 1967).

In what follows, the memory system will be assumed to be divided into three components: a sensory register (SR) which receives information from the sense organs; a short-term store (STS) which may temporarily hold information that has been passed to it, either from the SR or from the third component of the system, the long-term store (LTS). The LTS represents permanent memory, and it is only here that information[3]

[3]In this paper the term "information" is used to refer to codes, mnemonics, images, or other material that the subject places in memory and that can help him to generate a response; we will not use the term in its formal information-theoretic sense.

may be retained for an extended period of time. All three of these stores are capable of retaining information received from any of the sense modalities. Since the experiments that will be discussed in this paper have used verbal material exclusively, no attempt will be made to consider memory other than of a linguistic nature. This restriction does not represent a limitation of the theory, since the system can accommodate other sorts of material (see Atkinson and Shiffrin, 1968a, for a more complete discussion).

At the outset, it is important to make a distinction between two aspects of the proposed memory system. On the one hand, there are certain fixed structural features of the system that are invariant and cannot be modified by the subject. On the other hand, the operation of the system is determined by a set of control processes that may change from one point in time to another. Thus, for example, information that is transferred from the SR to LTS must pass through STS since the functional connections between the three states are structural aspects of the system. The way in which STS is used to make this transfer, however, is a control process selected by the subject that can be quite different in nature from one task to the next. In one task the subject may use STS to rehearse several items simultaneously in order to maintain them over a short retention interval, whereas in another task each item may be studied and coded individually in an attempt to form a mental image for longterm storage. We shall return to an example in which different uses of STS are illustrated after a brief description of the components and control processes of the system.

The interconnections between the three stores are illustrated in Fig. 4-1. New information can enter the system only via the SR. In order to be retained, it must be passed from there to STS. It is in this store that most processing of information takes place. The STS, therefore, receives input not only from the SR but also from LTS. Information may be transferred from LTS to STS, for example, during recall, during the formation of associations while coding an item, or during the comparisons of one event with the memory of another. Finally, information which is to be permanently stored in LTS is "copied" into it from STS. Notice that the transfer of information from one store to another is a nondestructive process; that is, the information in the original store is not lost as a result of a transfer per se.

In the case of visual input[4], the information entered into the SR usually takes the form of a fairly complete image of the observed scene which will decay in a matter of a few hundred milliseconds. The control processes at this level are concerned primarily with the selection of ma-

[4]The properties of the SR are best known for visual input; for some information on other modalities, however, see Crowder and Morton, 1969; Hill and Bliss, 1968.

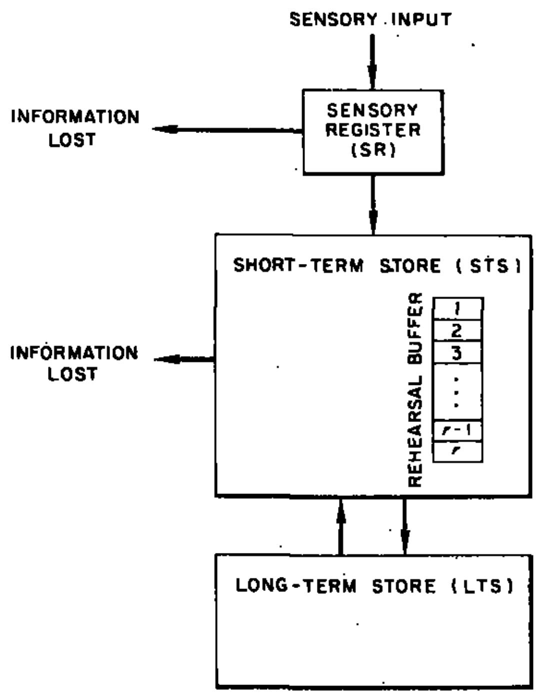


FIGURE 4-1. Structure of the memory system.

terial for transfer to STS. Much more information is present in the SR than it is possible to transfer to STS. For example, partial report studies of visual memory (Sperling, 1960) show that subjects are able to recall correctly one line of a tachistoscopically presented 3 × 4 array of letters if they are instructed which line to remember immediately after presentation. If the recall instruction is delayed by more than a tenth of a second, the number of letters that are correctly recalled drops sharply, indicating that information originally present in the SR was lost before it could be transferred to STS.

Information entered in STS will also decay, but at a slower rate than in the SR. The measurement of this decay is complicated by the fact that the subject is able to retain information in STS almost indefinitely by rehearsal. Experiments (e.g., Peterson and Peterson, 1959) which attempt to prevent rehearsal have generally indicated that, without rehearsal, information in STS decays with a half-life on the order of ten to fifteen seconds, the exact rate being highly dependent on the interpolated activity (Spring, 1968).

Control processes associated with STS may be grouped into three classes. The first of these classes is associated with the search for information in STS and its retrieval. There is evidence that the storage of information in STS is structured, hence that the use of a particular search strategy may lead to more or less rapid recovery of certain aspects of the data (Murdock, 1967; Sternberg, 1966). These search processes do not play an important role in experiments of the type that we shall be considering in this paper, so will not concern us further.

The second class of control processes in STS is far more important in the typical learning experiment. Processes of this type involve the rehearsal of items in STS in order to circumvent their decay. As long as information is rehearsed in STS it is preserved, but it begins to decay as soon as rehearsal ceases. In order to formalize this rehearsal process, it is assumed that the subject sets up a buffer in STS that can hold a fixed number, *r*, of items (see Fig. 4-1). This buffer is not a structural feature of the system, but is set up by the subject when required. The size of this buffer, when it exists, will depend both on the nature of the material that is being rehearsed and on the learning strategy that the subject is currently employing. It is not necessary that every item which enters STS be incorporated into the rehearsal buffer. The decision as to whether an item is to be entered into the buffer is another control process and depends on, among other things, the nature of the item and on the current contents of the buffer. Since the buffer is of fixed capacity, when an item is entered another must be deleted. The probability that a particular item in the buffer is forced out depends on such factors as the age of the item, the ease with which it can be rehearsed, and so forth (Brelsford and Atkinson, 1968). Once an item has been deleted from the buffer it undergoes rapid decay in STS.

The third important class of STS control processes are those associated with the transfer of information to LTS. In general, whenever information is in STS, some of it will be transferred to LTS. What is transferred, however, may vary greatly, both in the quantity and the quality of the resultant representation in LTS. If the major portion of the subject's effort is devoted to rehearsal in STS, relatively little information will be transferred to LTS, whereas if he attempts to develop appropriate ways of organizing and encoding the material, a great deal may be transferred. For example, in the learning of paired-associates, long-term performance is greatly improved if the subject searches for some word or phrase that will mediate between the stimulus and the response rather than simply rehearsing the item (Montague, Adams, and Kiess, 1966). Of course, the reduced rate of transfer to LTS as the result of the generation of a rehearsal buffer is frequently offset by the greater length of time which the information will reside in STS and hence be available for transfer to LTS. The size of the buffer can also affect the rate at which information is transferred in another way. All of the items in STS at any one time are, in a sense, competing for transfer to LTS. Thus, when the buffer is large, the amount of information transferred to LTS about each item is proportionally smaller.

In the view of this theory, information that is stored in LTS is not subject to decay. Information, once stored, remains in LTS indefinitely. This does not imply, however, that this information will always be immediately available for recall. It is essential here to distinguish between the storage of information in LTS and its retrieval. Information which has been stored at one time may fail to be retrieved at a later time either because the strategy which the subject employed to locate the information is inadequate, or because later learning may have resulted in the storage of additional information that was sufficiently similar to that stored about the item in question as to render the original information, when recovered, insufficient for the generation of a correct response. In general, the control processes which are associated with LTS are involved with storage and with the determination of appropriate search routines. These will not be important in the discussion of reinforcement to follow, so the reader is again referred to the papers by Atkinson and Shiffrin (1968a,b) and Shiffrin and Atkinson (1969).

In the remainder of this section, an unpublished study run by Geoffrey Loftus at Stanford University will be described. We have three reasons for presenting this experiment. First, it will illustrate the continuous paired-associate task that has been used in much of the experimental work to be considered later in this paper. Second, it will extend our discussion of the memory system, in particular indicating how it can be given an explicit quantitative formulation. Finally, the experiment will provide an illustration of the way in which control processes in STS are affected by the nature of the task.

In this experiment, subjects were required to keep track of a randomly changing response paired with each of nine different stimuli. To be more specific, the task proceeded as follows: At the start of the experiment each of the nine stimuli (which were the digits 1 through 9) was paired with a randomly selected letter from the alphabet. After these initial presentations the experiment proper began. At the start of each trial a randomly chosen stimulus was presented to the subject, and he was required to make a response indicating which letter had last been paired with it. As soon as the response had been made, the same stimulus was presented for study paired with a new response chosen at random from the twenty-five letters not just paired with the stimulus. The subject had been instructed to forget the old stimulus-response pairing and to remember only the new one. After a brief study period this pair disappeared and the next trial was started. In this manner three hundred trials could be presented during a session lasting about an hour.

The motivation for Loftus' experiment was to examine how the type of test employed to measure retention would affect the strategy used by the subject to store information. In particular, strategies were to be examined when the subject knew that he was to be tested using a recognition procedure, when he knew that a recall procedure was to be used, and when he had no information about the type of test. There were, thus, three experimental conditions, only one of which was used during a single session: (1) Items were tested by a recognition procedure; that is, at test a stimulus was presented along with a letter that was either the correct response or another randomly chosen from the remainder of the alphabet. The subject made his choice by striking either a key marked "YES" or a key marked "NO" to indicate whether or not he thought that the letter was indeed the one last paired with the stimulus. This condition will be referred to as the *recognition* condition. (2) Items were tested by a recall procedure; that is, a stimulus was presented alone for test and the subject was instructed to strike a key indicating which of the twenty-six letters of the alphabet he thought was correct. This condition will be referred to as the *recall* condition. (3) On each trial the choice of whether to use a recognition or a recall test was made randomly with equal probability. The data from this *mixed* condition must, therefore, be analyzed in two parts, according to which type of retention test was used. Unlike the other two conditions, when subjects were serving in the mixed condition, they were unable to tell at the time of study how that item would be tested.

Eight college students served in this experiment, each running for a total of sixteen daily sessions. In each session one of the three conditions was used. In order to allow subjects to become familiar with the apparatus and with the nature of the test procedures, the first session was run in the mixed condition and the data collected were excluded from analysis. During the remainder of the experiment each subject served in each condition for a total of five sessions. To avoid warmup effects during the later sessions, the first twenty-five trials of each session were also eliminated. The resulting data consists of 1,375 trials for each condition and each subject. The experiment was controlled by a modified PDP-1 computer which was operated on a time-sharing basis to drive eight KSR-33 teletypes, one for each of the subjects. These teletypes were used to present the material and to receive responses. The output from each teletype was masked so that only a single line of typed material was visible to the subject. This allowed control of the duration of the exposure and prevented the subject from looking back to the results of earlier trials.

Since the stimulus that was presented on a trial was chosen randomly, the number of trials that intervened between the study of a particular stimulus-response pair and its subsequent test was given by a geometric distribution with parameter equal to the reciprocal of the number of stimuli, in this case 1/9. The data which were collected, therefore, can be summarized by plotting the proportion of correct responses as a function of the number of trials that intervened between study and test. We shall refer to the number of intervening trials as the *lag* of the test for that item. In Fig. 4-2 the proportion of correct responses at a given lag is plotted for each of the conditions. There are over one thousand observations at lag zero for the recall and recognition groups and about half that many for the two curves from the mixed condition. The number of observations falls with increasing lag according to the geometric distribution mentioned above; thus there were only about two hundred observations for each condition by lag 14. Beyond this lag, therefore, the lag curves begin to show considerable instability and have not been plotted. The recognition data

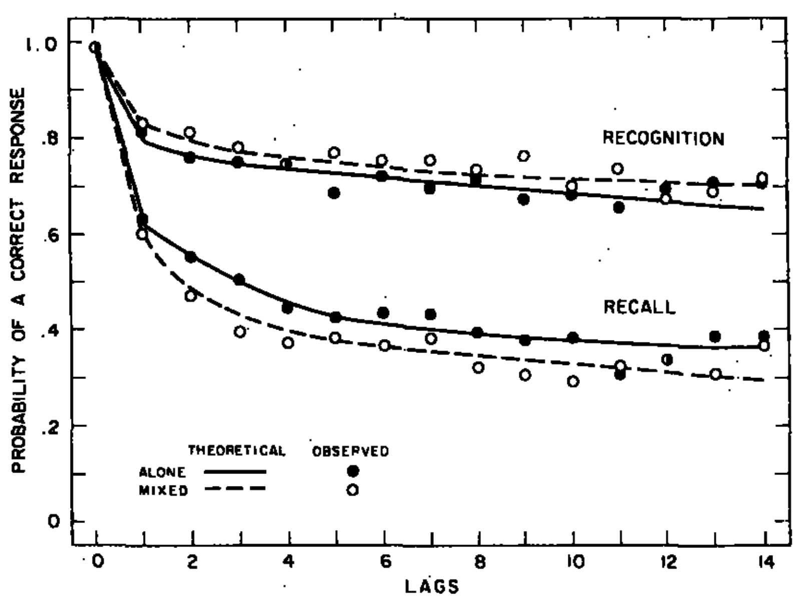


FIGURE 4-2. Probability of a correct response as a function of the lag between study and test for different retention-test conditions.

may be separated into two subsets, depending upon whether the pair presented to the subject for identification was actually correct or incorrect. In Fig. 4-3 lag curves reflecting this distinction are plotted: The upper curves show the probability of a hit (i.e., of a correct identification of a true pair) while the lower curves show the false alarms (i.e., the incorrect designation of a false pair as correct). These two functions were used in the analysis of the recognition data rather than the probability of a correct response.

The lag curves of Figs. 4-2 and 4-3 show a consistent difference between the mixed condition and the two homogeneous conditions. When serving in the recall condition, subjects were able to perform better than in the mixed condition. On the other hand, a greater proportion of the items were correctly recognized in the mixed condition than in the recognition condition. This result is also apparent in the proportion of hits and, to a lesser extent, of false alarms.

In order to interpret these results in terms of the memory system previously discussed, the assumptions of the theory must be given in a more explicit form (for a more detailed discussion of these assumptions and their implications, see Freund, Loftus, and Atkinson, 1969). The first step

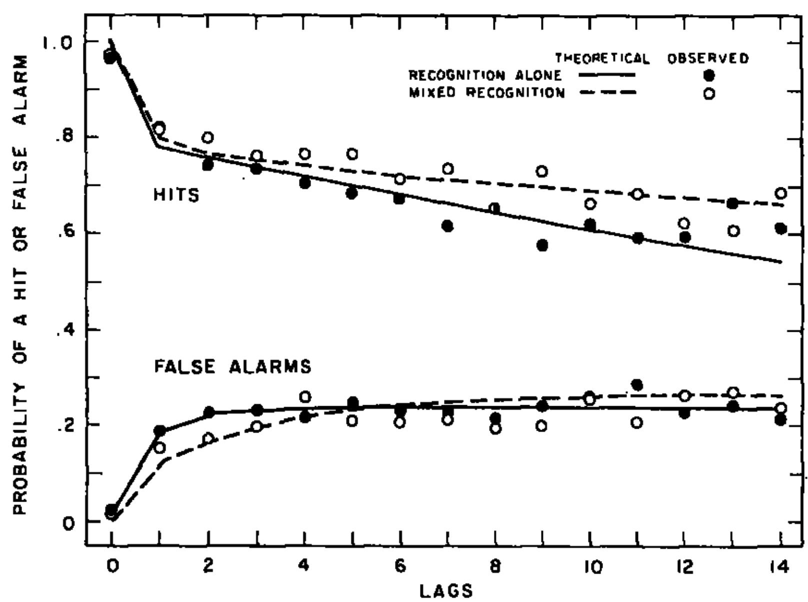


FIGURE 4-3. Probability of a hit and false alarm as a function of the lag between study and test.

is to clarify the conditions under which a new stimulus-response pair will enter a rehearsal buffer in STS. Whenever a stimulus is presented for study, there is a possibility that it will already be in the buffer, although the response that is paired with it will now be incorrect. If this happens, it is assumed that the new pairing invariably replaces the old pairing in the buffer. In the case where the stimulus that is presented for study is not represented in the buffer, we assume that entry is not assured, but takes place with probability α. The value of the parameter α is not known in advance and will need to be estimated from the data. If the new item enters the buffer, another item must be removed so that the buffer size remains constant at r items. As mentioned above, the choice of which item to delete from the buffer depends on many factors, but for this analysis it is sufficient to assume that it is random, with each item having the same probability of being knocked out.

The second set of assumptions that are required to make explicit predictions from the theory involves the transfer of information from STS to LTS. Since every item that is presented enters STS (although it does not necessarily enter the buffer), there will be some minimum amount of information about it transferred to LTS. This quantity of information will be denoted by θ'. If the item is also included in the buffer, it will reside in STS for a longer period of time, and hence more information about it will be transferred. In particular, it will be assumed that for each trial that passes, an additional amount of information, θ, will be transferred.[5] Thus, for an item which enters the buffer and resides in it for j trials, the amount of information in LTS will be $\theta' + j\,\theta$. For simplicity we identify the two transfer parameters θ' and θ so that the information transferred will be $(j + 1)\,\theta$.

Information once stored in LTS is postulated to remain there indefinitely. Nevertheless, with the passage of time, other information may also be transferred to LTS which makes the original information less easy to

[5]The model that is represented by this assumption may be contrasted with a "single pulse" model in which rehearsal in STS does not induce additional information to be transferred to LTS, that is, in which $\theta = 0$ but $\theta' > 0$ (Atkinson, Brelsford, and Shiffrin, 1967, Appendix). Evidence for the continual transfer assumption that we have used is provided by a free-recall experiment run by Dewey Rundus at Stanford University (Rundus and Atkinson, 1969). In learning the list of items to be recalled, subjects were instructed to rehearse out loud as the study list was being presented by the experimenter. This rehearsal was tape-recorded, and the set under rehearsal after the presentation of each new item could be precisely determined. Under these conditions the probability of correctly recalling an item when tested was a sharply increasing function of the number of times that it was in the rehearsal buffer: Items that were in the buffer for a single time period were correctly recalled only 12 per cent of the time, while items that were rehearsed for nine or more times were almost always given correctly.

retrieve or which renders it ambiguous once retrieved. To quantify this decrement we assume that *retrievable* information decreases by a proportion $1 - \tau$ for every trial which passes after the item has left STS $(0 < \tau \leq 1)$.[6] In summary, the amount of information which will be retrievable from LTS for an item that remained in the buffer for j trials and was tested at a lag of i trials $(i \geq j)$ is $(j + 1)\,\theta\tau^{i-j}$.

The final class of assumptions specifies the relationship between information in LTS and the production of an appropriate response. There are three cases to consider here, depending on the disposition of the item in STS. The first of these is the case where the test is at a lag of zero. It is assumed here that the correct response is always available in STS regardless of whether the item was entered into the rehearsal buffer or not. No error is made. Similarly, when the lag is greater than zero but the item has been entered into the buffer and is still resident in it, a correct response will be made with probability one. Only in the third case, when the item is not in STS and must be retrieved from LTS, is an error possible. The probability that a correct response is produced here will depend upon the amount of information transferred to LTS. There are a number of ways in which this correspondence can be made; in the analysis of the experiment considered here, a postulate based on signal-detection theory was used. This equated the sensitivity parameter, d', with the amount of retrievable information, i.e.,

$$d'_{ij} = (j + 1)\,\theta\tau^{i-j}.$$

For the recall data, this value can be converted to the probability of an incorrect response (Elliot, 1964) which we shall denote by η_{ij}. For the recognition data, the results must be analyzed in terms of hits and false alarms, requiring the introduction of a bias parameter, c, associated with the subject's tendency to respond "YES."

[6]In a more precise model of memory the decay of information in STS would be represented by the same sort of exponential process that we have used here to describe the deterioration of information in LTS. This loss of information would be through actual decay, however, rather than through problems of retrieval that have been postulated for LTS. Formally, parameters θ'' and τ'' would be required, the first representing the amount of information available in STS at the time when an item is knocked out of the buffer, the second representing the rate of decay of this information in STS. The amount of information retrievable from both STS and LTS would, therefore, be $(\theta' + j\theta)\,\tau^{i-j} + \theta''\tau''^{i-j}$. The original amount of information in STS would be greater than that in LTS ($\theta'' > \theta$ or θ'), but its rate of decay would be more rapid ($\tau'' > \tau$) so that the short-term contribution would become negligible while the contribution of LTS was still large. For the purposes of the analysis at hand, however, we can assume that information in LTS becomes unavailable so much more slowly than in STS, that the short-term decay factors may be ignored without changing the quality of the predictions.

The final step in the analysis involves the calculation of the actual probabilities of correct or error responses. From the assumptions about the probability that an item enters the buffer and that it is later forced out, we can calculate the probability that an item resides in the buffer for exactly j trials given that it is tested at a lag greater than j. This probability will be denoted as β_j. Since errors may occur only when the item is not in the buffer (i.e., only when it has resided in the buffer for a number of trials less than the lag), the net probability of an error is equal to the probability that an item remains in the buffer j trials multiplied by the probability of an error given this number of trials in the buffer, these terms summed over values of j less than or equal to i. Hence, the probability of an error at lag i is

$$P(E_i) = \sum_{j=0}^{i} \beta_j \eta_{ij},$$

where the case of $j = 0$ is used to indicate that the item did not enter the buffer. The derivation of the hit and false-alarm functions follow very much the same pattern.

The predictions of the theory, therefore, depend on the integer-valued parameter r and on the four real-valued parameters α, Θ, τ, and c. In order to estimate these parameters, a minimum chi-square procedure was used. For the recall condition, the observed frequencies of correct responses and of errors were compared to their predicted values with a standard Pearson chi-square. Because the probabilities of correct responses are not independent at different lags, the result of this calculation is not assured of being distributed as a true chi-square. Nevertheless, it should be approximately correct and in any case should be nearly monotone in goodness of fit. The set of parameters that minimize the chi-square will, therefore, be a good estimate of the true parameter values. In order to evaluate approximately how well a particular parameter set fits the data, the resultant "chi-square" can be compared with a true chi-square distribution. For this comparison, each of the fourteen points on the lag curve will contribute a single degree of freedom to the chi-square. Subtracting one degree of freedom for each of the four parameters estimated (performance in the recall conditions does not depend upon c) the total number of degrees of freedom is $14 - 4 = 10$. In the case of the recognition condition, the data consist of two functions, the hits and the false alarms. By fitting both of these functions simultaneously, the number of degrees of freedom in the initial sum is doubled. Since in this case five parameters are to be estimated, a total of $2 \times 14 - 5 = 23$ degrees of freedom are available. Finally, for the mixed condition, minimization must be carried out simultaneously over the hits and false alarms for the recognition data and the number of correct responses for the recall data. There are, then, thirty-seven degrees of freedom in this chi-square.

TABLE 4-1

ESTIMATES OF MODEL PARAMETERS FOR PAIRED-ASSOCIATE ITEMS TESTED BY A RECOGNITION, A RECALL, OR A MIXED PROCEDURE

	Experimental condition		
	Recognition	*Mixed*	*Recall*
r	1	2	3
α	0.79	0.73	0.53
Θ	0.79	0.52	0.30
τ	0.95	0.97	0.99
c	0.71	0.62	*
χ^2	22.3	29.3	11.3
df	23	37	10

*The paramenter c was not required for this group.

The results of these estimations are shown in Table 4-1. It is first worth noting that the chi-squares are roughly on the same order as the number of degrees of freedom, and so in every case the fit is satisfactory. However, because the assumptions of the Pearson chi-square are not satisfied here, a comparison of the relative goodness of fit between the groups may not be made.

The values taken by the five parameters indicate the nature of the differences between conditions. The changes in all of the parameters are monotonic across the three conditions, with the mixed condition showing estimated parameters between those of the two unmixed conditions. The parameter c is not too useful here since it was estimated for only two of the conditions and since it does not differ much between them. The parameter that changes most dramatically is the size of the buffer, r. This parameter is estimated at 1 for the recognition condition, at 2 for the mixed condition, and at 3 for the recall condition. At the same time the probability that a new item enters the buffer, α, drops from 0.79 in the recognition condition to 0.53 in the recall condition. This difference in parameters implies that in the recognition condition subjects enter most items into the buffer, but hold them there for little more than a single trial, whereas in the recall condition almost half of the items fail to enter the buffer at all, although when they do enter, they tend to stay for a fairly long time. The mean number of trials that an item stays in the buffer, given that it is entered, is r/α, which is 1.3 trials for the recognition condition and 5.7 trials for the recall condition. At the same time, the amount of information about each item that is transferred into LTS on each trial, indicated by the value of Θ, is much larger for the recognition condition than for the recall condition.

These results may be interpreted as characterizing two alternative strategies that the subject can adopt to deal with the two testing procedures. When the recognition test is used, the quality of the information required to produce a correct response is fairly low. It would, for example, frequently be sufficient to code the response letter E simply as an early letter in the alphabet or as a vowel. In this condition the parameter estimates suggest that the subject chose to concentrate on each item when it was presented and to transfer as much information about it as possible to LTS. Although the quality of this representation was probably poor and became largely unavailable at long lags ($\tau = 0.95$, but, e.g., $\tau^{8} = 0.66$), it was frequently sufficient to determine a correct response. On the other hand, the recall condition required much more complete information. Apparently, in this condition the subjects tried to maintain some items in STS for a longer time, at the expense of other items. A strategy similar to that used for the recognition condition apparently transferred so little information to LTS as to be unable to support recall. The strategy employed, therefore, seems to be to use STS as much as possible for information storage (remember that more short lags are present than longer lags), even though this allowed information about each item to accumulate in LTS only slowly ($\Theta = 0.30$ compared to 0.79 for the recognition group). In order to do this, some incoming items had to be skipped almost entirely. In the mixed condition, subjects apparently were forced into an intermediate strategy, retaining items in STS for longer than they had in the recognition condition, but not for as long as in the recall condition. It is interesting to note that fewer errors were made on the recognition task in the mixed condition than in the recognition condition. Apparently, the strategy selected for the mixed condition actually was better on recognition tests than the strategy selected when the recognition task only was present. It seems that subjects do not always choose the set of control processes which produce the best performance.[7]

[7]The interpretation given to the above experiment is based in part on the parameter estimates presented in Table 4-1. It should be noted that the interpretation also depends on a detailed analysis of the sequential properties of the data that have not been described here. The reason is that such analyses are complex and require a lengthy description; further, analyses of this sort will be considered later in treating a similar experiment (pp 88–97).

Attention. It is difficult to consider the concept of reinforcement without at least attempting to relate it to attention.[8] The extent to which a particular event modifies a subject's later behavior is influenced by the attention he gives to that event as much as by any reward or punishment associated with it. Accordingly, before reinforcement is considered, we shall examine the ways in which attentional variables can be incorporated into the framework of our memory system. We assume that attentional variables affect this system in three different ways, associated with the input of information into the SR, STS, and LTS. In the next section, when considering reinforcement, our interpretation of it will be very similar to the third of these attentional processes: that associated with entry of information to LTS.

The first place where attention can affect information transfer is at the very outset, by selecting information for entry into the SR. The processes which determine this selection are, in general, gross behavioral ones, primarily involving the orientation of the subject toward sources of stimulation so that the appropriate sense organs are stimulated. Once the sense organs have been activated, however, we assume that the incident information will be transferred to the SR.

The attentional processes involved in the transfer of this information to STS are more complex. This transfer results in a great reduction in the amount of information that is processed, since only information of importance to the subject is entered into STS. Such information may roughly be grouped into three classes which we associate with three different types of transfer control processes. The first class of information transferred to STS relates directly to the task with which the subject is currently involved. Thus, for example, in reading this text, one more or less automatically transfers information about the next words into STS (note, however, that the eye-movements involved in scanning the page are an attentional process of the first type). To account for this transfer, it will be assumed that the presence of information of a particular sort in STS will induce transfer of any similar information in the SR to take place. It is immaterial whether the control processes involved here are thought of as comparing the contents of the SR to STS, or as reaching out from STS and tracking a particular part of the SR. In any case, these control processes allow the system to track activity in the environment as long as information about it is maintained in STS. The second class of information transferred requires a somewhat more elaborate set of control processes. It is postulated that all information entered into the SR is rapidly analyzed

[8] See Guthrie (1959) for an interesting discussion of this point.

and, as part of this analysis, a reference is made to LTS. At this stage, the primary result of this reference is the retrieval of a quantity, the *pertinence* associated with the information (Norman, 1968). For the purposes of this discussion, the pertinence may be thought of as a scalar quantity, with the property that information which has a high pertinence is likely to be selected for transfer to STS and information which has a low pertinence is likely to be allowed to decay without attention. The value that is taken by the pertinence function will depend on many different variables. The recency of a reference to the information in LTS and the frequency with which the information has been referenced, for example, are two such variables. The reference to LTS and the transfer to STS take place only after the information in the SR has been analyzed at a fairly high level. If anything is entered into STS as a result of these attentional processes, it will be far more complicated than a sensory image and will include some of the information recovered from LTS, for example, its context and several associations to it. The last class of information which may be transferred from the SR to STS concerns sudden changes in the environment. It is postulated that whenever there is a sharp discontinuity in the contents of the SR that is not correlated with an observing response or other subject-induced activity, there is a tendency for the new material in the SR to be transferred to STS. It is worth noting that these three classes of processes are competing with each other for the limited processing capacity available in STS, as well as with information that is being transferred from LTS and information that is being maintained in STS. What actually will be entered depends on the relative demands of all these sources of input, rather than on the magnitude of any one request.

The third place where attention influences the transfer of information is in the link between STS and LTS. It is clear that we remember a great deal about some aspects of the environment and very little about others, even when we have "attended" to all of them. In interpreting such effects it is not necessary to add anything to the collection of control processes that have already been introduced. In the previous section we noted that the transfer to LTS was influenced by any of a number of control processes acting on STS. The number of items in STS, the formation of a rehearsal buffer, or the retrieval of information from LTS to form mnemonics are examples of these processes. We shall not dwell on these attentional processes here, since they will be discussed in the next section.

The concept of reinforcement. In the preceding two sections a theory of memory and attention has been outlined that we believe can account for most of the results from simple verbal-learning experiments. In this section an attempt will be made to discuss reinforcement in the framework of this system. We do not think that a single formulation can explain the variety of reinforcement effects that have been demonstrated with human subjects. Rather, it appears that the major determinants of learning are the memory and attention processes, and that the concept of reinforcement may best be understood in terms of their action. In several of the applications to be discussed in the second part of this report, results will be presented where the reinforcement effects appear at first glance to be quite complicated. When these effects are analyzed in terms of the theory, however, their basis will be seen to be relatively simple. The memory and attentional processes available to the subject provide bounds, often quite strict, that limit the set of control processes that can be used, and thereby constrain the action of reinforcement.

In many ways our interpretation of reinforcement is quite similar to the ideas of attention that were discussed in the preceding section. Transfer of information to LTS takes place only while that information is resident in STS. Thus, if learning is to take place, the appropriate information must be maintained in STS for some period of time. As indicated before, however, STS is a system of limited capacity, and many potential sources of information are competing for access to it. At the same time that an item is being studied for later recall, processing space in STS is also demanded by incoming stimuli and by other items already in STS. The extent to which information about the item is successfully processed depends on the limitations imposed by the task and on the strategy selected by the subject.

The data collected in an experiment may appear to be unduly complicated for another reason. The system of memory has two distinct ways in which information about an item may be stored. An improvement in performance as a result of a study trial may be brought about either because information is temporarily maintained in STS or because it is permanently stored in LTS. The relative importance of these two stores will depend on many factors, such as the nature of the task, the presence or absence of competing stimulation, and the length of time between study and test. The operation of reinforcement will have an effect on both of these processes; that is, feedback or payoff may lead the subject both to retain information in STS and to try to transfer it efficiently to LTS. Although the term reinforcement typically is used to refer to processes that have an effect on the permanent storage of information, in many experiments these long-term effects can become confused with those due to STS. The long-term and short-term effects may be very different from each other. In the next section, for example, we shall consider an experiment in which the effects of a series of similar stimuli on the storage of information in LTS agree with predictions from classical interference theory, whereas the effect on the contents of STS is exactly the opposite. The overall behavior is, of course, a mixture of long- and short-term effects and thus, at first analysis, appears to show inconsistencies. In short, we do not feel that it is possible to study reinforcement variables without first making a careful analysis of the role of the two types of memory in the learning situation.

There are actually at least three sets of control processes by which information can be maintained in memory for later use. If the information is to be used immediately and then can be discarded, the subject may choose to simply maintain as much of it as possible in STS via rehearsal without any attempt to transfer it to LTS. With such a strategy the subject will be highly accurate at short lags, but performance will drop rapidly to chance thereafter. The second type of strategy also involves maintenance of information in STS via rehearsal, but this time in lesser quantity so that an attempt can be made to transfer it to LTS. Again, performance will be good at short lags, but now items tested at long delays will not experience as large a drop in performance. Finally, the subject may attempt to code the information and store it in LTS as it comes along without maintaining it in STS for any length of time. This set of control processes usually involves the retrieval of information from LTS to help generate a more robust image for permanent storage, usually by forming associations or by the use of mnemonic devices. The choice of which of these control processes to use is usually not freely available to the subject. The nature of the material that is presented frequently restricts the possibilities or even dictates exactly the method that must be used. The dynamics of the information processing that goes on in the three cases is different, however, and so the effect of an external manipulation will depend on the particular control processes that are used. In a later section on reinforcement magnitude, a case will be seen where a seemingly minor change in the stimuli led to a change in study procedure, which in turn resulted in vastly different reinforcement effects. An analysis of the information-transfer aspects of the situation is necessary before the role of reinforcement can be understood.

In spite of the restrictions that have been set forth in the previous paragraph, we shall now consider a general description of the reinforcement process. This formulation should not be thought of as an exact statement of the action of reinforcement, but as an outline which is frequently modified in its specifics. This description is, basically, an expectancy interpretation of reinforcement, and as such is in the tradition of the ideas set forth by Tolman (1932) and by Brunswik (Tolman and Brunswik, 1935). Essentially, it consists of two components: first, the formation of a prediction (and possibly the production of a response) based on the stimulus input and on correlated information retrieved from memory, and second, the comparison of this prediction with subsequent events. It is the result of this comparison that determines whether information about the episode will or will not be transferred to LTS.

As noted in the section on attention, the transfer of information about an external event to STS involves more than simply a transfer from the SR to STS. In particular, a reference to LTS is required in order to generate a pertinence measure, and some of the recovered information will be entered into STS along with information from the SR. This information, along with other information that may be retrieved later from LTS, is used by the subject to select a response if one is necessary. In addition, this information allows the subject to generate an expectation or prediction about the events that will follow the stimulus. Any response that is required is based on this prediction, but the prediction usually is more elaborate than may be inferred from the observable response. When the outcome event in question occurs, it is compared with this prediction. The extent to which the outcome fails to agree with the prediction determines the degree and nature of the study the item receives. Usually, large discrepancies between the prediction and the outcome dispose the subject to apply control processes that maintain the relevant information in STS and induce the transfer of information to LTS. The information which is transferred is primarily associated with those components of the prediction that were most deviant from the actual outcome. The result is to reduce the disparity between the outcome and information now stored in LTS, so that if the same stimulus and outcome were to be presented again, the discrepancy would be smaller than the original one.[9]

This special analysis simplifies considerably the factors that are involved in causing information to be maintained in STS. It is important to realize that STS is a system of limited capacity and that many potential sources of information are competing for access to it. At the same time that a comparison between a prediction and an outcome indicates a discrepancy, the processing capabilities of STS will also be demanded by external inputs and by other information that is already resident in STS. Whether the item in question will actually receive sufficient processing in STS to have an effect on later performance will depend upon the task in progress, the nature of the competing items, and any control processes

[9]The above hypothesis is similar to several other theories that have been proposed. The notion that the condition under which learning takes place involves a discrepancy between a prediction and an outcome is quite close to the expectancy hypothesis developed by Kamin (1969) and by Rescorla (1969). In the restriction of the stored information to that necessary to eliminate an observed discrepancy, our theory is similar to the discrimination net models of Feigenbaum and Simon (Feigenbaum, 1963) and Hintzman (1968). In this respect it also bears a resemblance to dissonance theory (Festinger, 1957; Lawrence and Festinger, 1962).

which may predispose the system to treat information of one type and not of another. This dynamic aspect of short-term processing is responsible for many of the effects of reinforcement, and we shall return to it in several of the applications that will be considered in the remainder of this paper.

Experimental Results

In this section the results of a number of experiments are considered in order to help clarify the role of the various stores and control processes and illustrate how reinforcement variables (e.g., the magnitude of reinforcement, the schedule of reinforcement, or the delay of its presentation) may be interpreted. In the original reports where these experiments were first described, they were given some form of quantitative analysis in terms of the theory. The details of these analyses can be found in the reference articles, so our discussion will be of a more qualitative nature. We hope that this simplification will allow us to consider the problems of reinforcement without becoming involved in questions of mathematical notation and proof.

Number of reinforcements and their presentation schedule. The first experiment is a fairly direct application of the theory to paired-associate learning (see Brelsford, Shiffrin, and Atkinson, 1968, for a more complete treatment). It illustrates the way in which a series of reinforcements can act to build up the strength of a representation in LTS through the successive storage of information. Basically, the same continuous paired-associate task that has already been described in connection with the Loftus experiment is employed, although with several modifications. A new set of eight stimuli (random two-digit numbers) were chosen at the start of each session and were used throughout the session. As in the Loftus experiment, the responses were letters of the alphabet. Each trial of the experiment began with the presentation of a stimulus to which the subject had been instructed to respond with the most recently paired letter. This stimulus was chosen randomly from the set of eight stimuli, so the lags between study and test were again distributed geometrically with parameter 1/8. Following his response, the subject was given three seconds to study the stimulus paired with a response. This ended the trial. Unlike the Loftus experiment, the study phase of the trial did not always involve pairing a new response with the stimulus. A stimulus-response pair might be given one, two, three, or four reinforcements, the probabilities of these frequencies being 0.3, 0.2, 0.4, and 0.1 respectively. Thus, a stimulus selected for two reinforcements would be studied with the same response following the first test, but after the second test a new response would be introduced. This procedure continued for 220 trials per session. Each subject was run for at least ten sessions.

As in the previous experiment, the principal finding can be expressed in the form of lag curves (Fig. 4-4). Separate curves are presented showing the probability of a correct response, depending upon the number of prior reinforcements. Hence, there is a lag curve for stimulus-response pairs tested after one, two, and three reinforcements. By the nature of the presentation schedule, the number of observations at each point declines with increasing lag, and also with increasing number of reinforcements. Since at the time a subject was tested on an item, he had no way of knowing whether that item would be studied again, the first test of every item could be used in plotting the lag curve for one reinforcement. Similarly 70 per cent of the items received two or more reinforcements and therefore contributed to the second lag curve. Only in the case of the fourth reinforcements (which occurred for only 10 per cent of the items) were the frequencies too small to permit stable curves to be plotted. The three curves in Fig. 4-4 show a resemblance in form to the lag curves

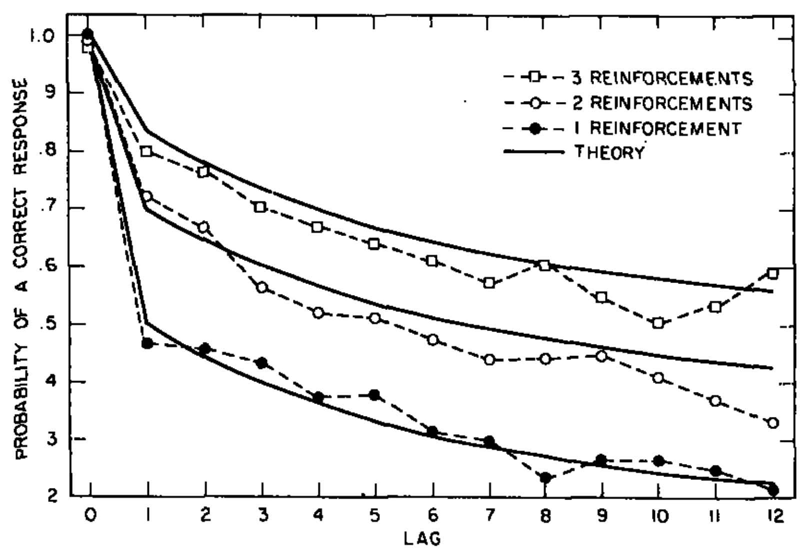


FIGURE 4-4. Observed and predicted probabilities of a correct response as a function of lag for items tested following their first, second, or third reinforcement.

obtained in Loftus' experiment. In particular, the curve for one reinforcement is quite similar to the comparable curve for the Loftus recall group. The curves in Fig. 4-4 also indicate that the proportion of errors at a given lag decreased as more reinforcements were given.

In order to account for the effects of multiple reinforcements, only a few minor changes need be made in the model used to analyze Loftus' data. As before, it is assumed that if a stimulus is presented for study paired with a new response and the stimulus is one of the r-items currently in the rehearsal buffer, then the subject will simply replace the old response with the new one. Otherwise, no change is made in the contents of the buffer. The case of an item which is not in the buffer at the time of presentation is somewhat more complicated.[10] Whenever the stimulus for such an item is presented for test, the subject must retrieve information from LTS in order to make a response. Again we assume that the amount of available information can be represented as a d'-measure for that item. On the basis of this information, the subject generates a response, in this case his prediction about the outcome of the trial. Accordingly, we postulate that whenever the response is correct (indicating a good correspondence between the prediction and the outcome), the item will not receive additional study and hence will not be placed in the buffer. Whenever the correspondence is small (an error is made), the item will enter the buffer with probability α. The probability of failing to enter the buffer, $1 - \alpha$, represents the combined effects of the many sources of competition in STS that may take precedence over entry of an item; for example, the presence of a naturally compatible stimulus-response pair or of an easily rehearsable combination of items in the buffer. Once the item has entered the buffer, however, we assume that transfer to LTS takes place in the same manner as discussed before: For every trial in the buffer an amount of information θ is transferred to LTS. Every trial in which the item is absent from STS results in a proportion $1 - \tau$ of the information in LTS becoming unavailable for recovery and response production. Like the recall condition in the previous experiment, the predictions of the theory depend on the four parameters: r, α, θ, and τ. To make these estimations, the same type of pseudo-chi-square procedure employed in the Loftus study was used here, this time simultaneously on all three lag curves and also on the double lag curves presented in Fig. 4-5. From this

[10]The analysis used here is not quite identical to that used by Brelsford et al. (1968, p. 6), the principal change being in the mathematical form of the response-generation postulate. The quantitative predictions of the two formulations are virtually identical; the one that is presented here is more in line with our current thinking regarding reinforcement. In the version of the theory used by Brelsford et. al., the parameters have slightly different meanings, and hence their values cannot be directly compared with those estimated for the Loftus experiment.

minimization, a set of parameters was found which generated the predicted curves shown in Figs. 4-4 and 4-5 and in the subsequent figures. The estimated buffer size was $r = 3$.

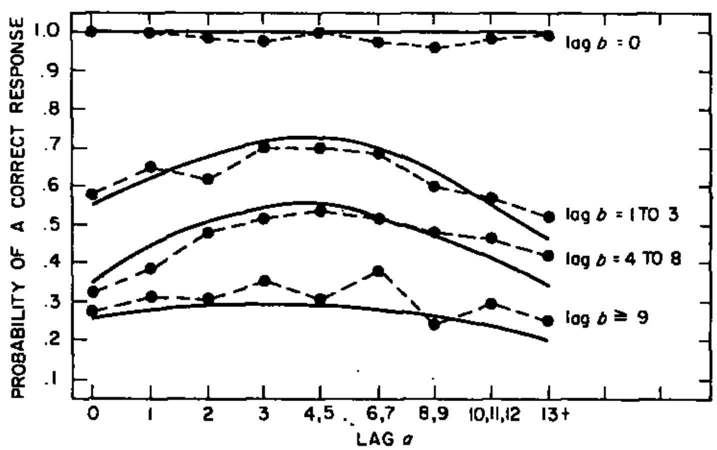
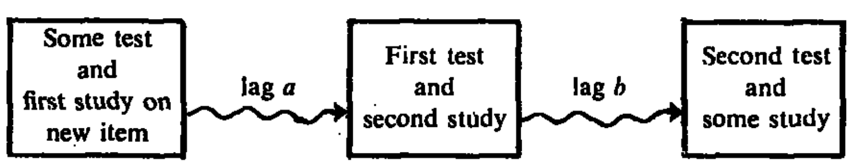


FIGURE 4-5. Observed and predicted probabilities of a correct response as a function of the spacing between the first and second reinforcement (lag a) and the lag between the second reinforcement and the final test (lag b).

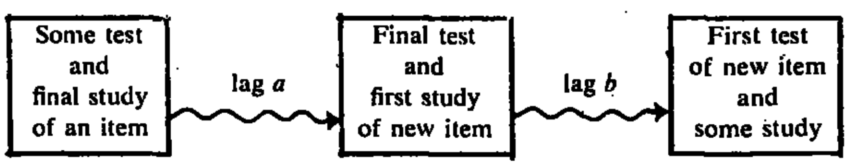
The lag curves of Fig. 4-4 give a good idea of the general rate of learning, but they are not the best way to look at the effects of reinforcement. These effects are better examined by looking at sequential properties of the data; that is, at the effects of one reinforcement on a later one. Accordingly, in the next few paragraphs we consider a number of different summaries of the data, and show how they are predicted by the theory.

The first set of results to be examined relates the lag between the first study and test of an item to the performance on the second test. In particular, the presentation of an item with two or more reinforcements can be represented as follows:



This describes a new pair that is studied, then first tested at lag a, is studied again, and next tested at lag b. We wish to look at the way in which results of the second test depend on lag a, with lag b held roughly constant. Plots of this relation are shown in Fig. 4-5. For lag $b > 0$ these curves are bow-shaped, with fewer correct responses when lag a is either small or large. As would be expected from the curves in Fig. 4-4, more errors are made when lag b is large than when it is small. It is relatively easy to see how these curves are predicted by the model. For small values of lag a, little information will be transferred to LTS during the interval between trials, so the primary effect of the first reinforcement is to increase the likelihood that the pair is in STS when the second reinforcement occurs. This will slightly increase the probability of a correct response, particularly at short lag b. For somewhat longer values of lag a, this effect is coupled with the transfer of a considerable amount of information into LTS before the second study. Thus a facilitative effect of the first reinforcement is expected even when the item has been deleted from the buffer before the second test. Finally, when lag a is very large, the item will almost certainly have departed from the buffer and much of the information that had been deposited in LTS will have become unavailable (in this experiment the estimate of τ was 0.82, so the retrievable information in LTS had a half-life of only about three trials).

In the preceding paragraph the effect of the lag between the first and second reinforcement of a stimulus-response pair was examined. In this paragraph we shall again consider the effects of the lag between two successive reinforcements involving the same stimulus; however, in this case the two presentations represent the last occurrence of one pairing and the first occurrence of a new pairing:



Here, a stimulus-response pair is given its last study and tested at lag a. A new response is then paired with the stimulus and is given its first test at lag b. The predictions for this case are somewhat surprising and are worth examining closely. If the item is not in the buffer at the end of lag a, it should have no effect on whether the new pairing is studied or not. If the previous stimulus-response pair is in the buffer, however, it should have a facilitative effect on the new learning, since the new item is now guaranteed to enter the buffer. In this case, the probability of a correct response on the new item should be relatively large. Unfortunately, the presence of the pair in the rehearsal buffer is not an observable event, but it is probabilistically related to the occurrence of an error and to lag a.

In particular, if an error was made on the final test of the old item, we know that it was not in the buffer, and therefore predict that the probability of a correct response on the new item, when tested later, will be independent of lag *a*. When a correct response is made on the old item, it may be in the buffer, and furthermore, it is more likely to be in the buffer if lag *a* is small. In this case, small values of lag *a* should be associated with fairly large probabilities of a correct response, and these probabilities should fall with increasing lag *a*. Note that this prediction is quite different from what would be predicted by interference theory, since it associates good performance on a transfer task with good performance on original learning.

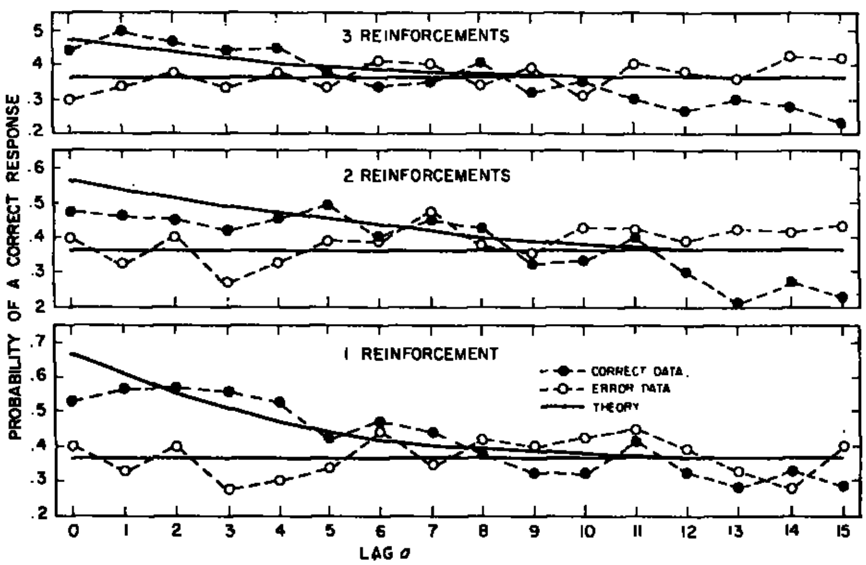


FIGURE 4-6. Observed and predicted probabilities of a correct response on the first test of an item as a function of the lag for the last item using that stimulus (lag *a*).

This prediction, however, seems to be well supported by the data as indicated by the functions plotted in Fig. 4-6. In this figure, unlike Fig. 4-5, the results have been averaged over all values of lag *b*. Three sets of curves have been plotted, depending upon whether the item given on trial $n + a + 1$ received its first, second, or third test. It is interesting to note that the magnitude of the difference between the correct and the error data declines as the number of prior reinforcements increases. This may be attributed to the fact that the facilitation is purely a result of study in STS, and that this study takes place only when the subject's prediction based on LTS information is incorrect. When several reinforcements have been given, there is a greater likelihood that the item will be correctly recovered from LTS, and hence that no rehearsal in STS will take place. Accordingly, the proportion of correct responses that occur because the item was maintained in STS decreases, and with it the size of the facilitation effect. It should also be noted that the probability of a correct response to the new item, conditional on a correct response to the old one, appears to fall systematically below the prediction when a long lag intervenes between the two study trials. This effect, which is exactly the opposite of the one observed at short lags, is evidence for the activity of more conventional interference processes in LTS. Items that are correctly recalled at long lags are likely to have been recovered from a good representation in LTS. Apparently this strong trace interferes with the establishment of a new trace based on the same stimulus. Additional evidence for these interference effects will be presented in Fig. 4-8.

The last two results to be considered involve the effects of a sequence of similar or dissimilar stimuli and provide further evidence for some of our postulates about study effects in STS. Consider a series of consecutive trials all involving the same stimulus, but in which the response paired with the stimulus on the final study trial is different from that on the immediately preceding trial. The theory predicts that the longer the string of presentations, the more likely it is that the final item when eventually tested will be correctly recalled. This is so because the probability that a pair containing the stimulus is in the rehearsal buffer increases with the sequence of zero-lag presentations. On each successive trial of this sequence, a pair containing the stimulus may be entered into the buffer if it is not already there, and if there are no competing items to force it out. The resulting effect is shown in Fig. 4-7. In this figure the probability of correctly recalling the last item of a series of trials all involving the same stimulus (averaged over all test lags) is plotted as a function of the length of the series. As expected, this is an increasing function, and falls quite close to the predicted function. Note that again this effect is quite the opposite of predictions from a traditional interference theory. Such a theory would predict that the repeated presentations would interfere proactively with the new pair and that this would decrease the probability of responding correctly to the transfer item. It is important to realize that these effects are the result of activity in STS and say nothing about the nature of interference in LTS. Indeed, the long-term effects appear to be the opposite of the short-term effects. Figure 4-8 shows the probability that, on the first trial of a new item, the response that had been correct

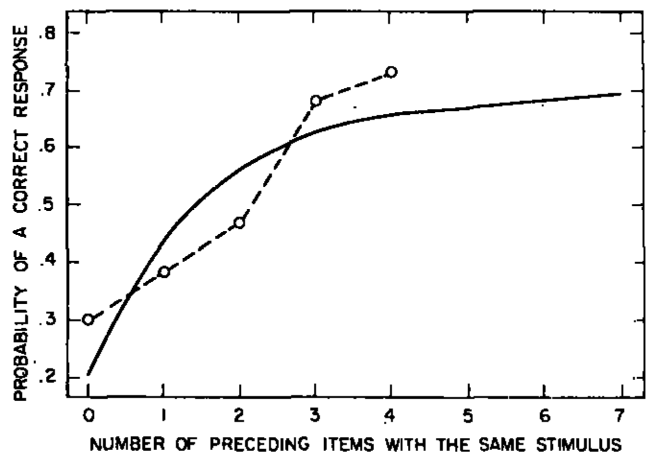


FIGURE 4-7. Observed and predicted probabilities of a correct response as a function of the number of consecutive preceding items using the same stimulus.

on the previous occurrence of the stimulus is given instead of the current correct response. The probability of these intrusion errors is plotted as a function of the lag at which the new item is tested (the three curves depend on the number of times that the previous pairing had been reinforced). Intrusion errors were more frequent when the previous item had been given several reinforcements than when it had received only a single reinforcement. The fact that the response is actually an error indicates

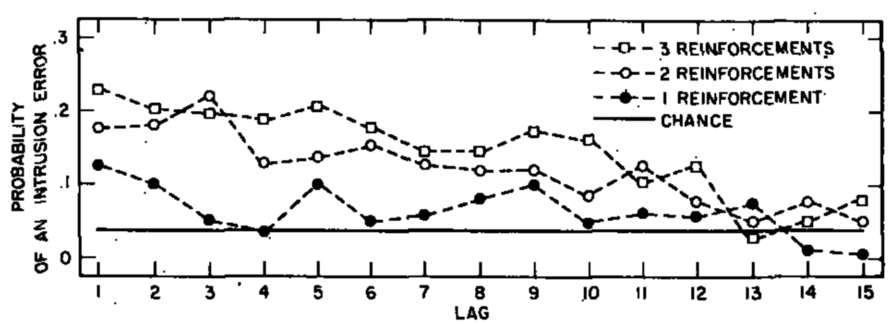


FIGURE 4-8. Probability that the correct response for the preceding item using a given stimulus will be made as an intrusion error to the present item.

that the item was not in the buffer at the time of test, hence that this more typical proactive effect is associated with long-term storage.

A series of consecutive trials using the same stimulus, as indicated in the preceding paragraph, tends to cause that stimulus to be entered into the rehearsal buffer, but will not create any further disruption of other items in the buffer. On the other hand, a series of items with different stimuli produces maximum disruption, since each of them will have some probability of being entered into the buffer. This effect is illustrated by the way that the items which intervene between study and test of a given item affect the probability of a correct response. In particular, suppose that the test of an item following its k^{th} study occurs at lag x. The case where all of the x intervening items involve the same stimulus and the case where they involve all different stimuli will be examined, with the prediction that the all-same condition will produce better performance than the all-different condition. For each of the three values of k, this prediction is supported (Fig. 4-9).

This experiment has illustrated the way in which the theory can be applied to show increases in LTS strength as a result of a series of reinforcements. It has also shown a simple way in which the correspondence

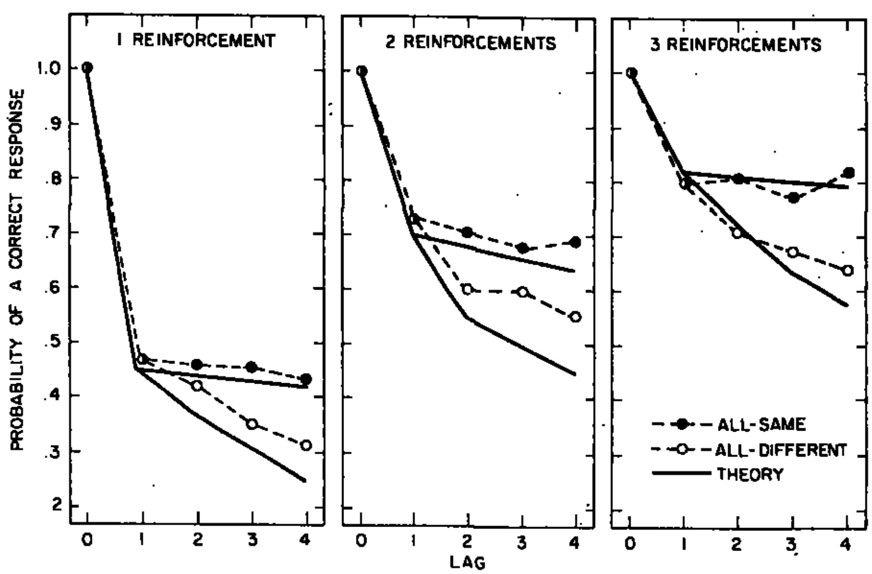


FIGURE 4-9. Observed and predicted probabilities of a correct response as a function of lag for the cases where the intervening stimuli are all identical or are all different.

between the subject's prediction and the outcome of a trial can determine rehearsal patterns. Finally, by considering the sequential properties presented in the last five figures, evidence has been given which supports our particular two-process formulation of memory.

Delay of reinforcement. The second experiment to be considered examines one of the most confusing issues in the area of human reinforcement: that of its delay. It appears that a delay in the feedback of information about a response can have many different effects. Some studies (Greenspoon and Foreman, 1956; Saltzman, 1951) have indicated that a delay will impair learning, others show no effect (Bilodeau and Ryan, 1960; Bourne, 1966; Hochman and Lipsitt, 1961), and still others appear to show a facilitative effect of delay (Buchwald, 1967, 1969; Kintsch and McCoy, 1964). We shall attempt to show that any of these effects can be accommodated by our analysis and shall discuss an experiment (Atkinson, 1969) in which all of these effects were obtained as the result of several fairly simple manipulations.

The basis of this experiment was a continuous paired-associate task similar to the one just described. The stimuli were randomly generated consonant trigrams and were paired with single-digit responses (digits 2 through 9). Every stimulus-response pair received between three and seven reinforcements, with each pair being equally likely to receive any number of reinforcements within this range. A stimulus was used only once during the course of the experiment; that is, a stimulus trigram would receive several study and test trials with a particular response number, and then would never be used again. The major difference between the presentation schedule in this experiment and those discussed earlier concerned the lag structure. Sixteen different stimuli were active at any time. The stimulus that was presented, however, was not chosen at random from this set, but only from the six stimuli that had not been presented on the previous ten trials. Thus, the minimum possible test lag was ten and the mean lag was fifteen items.

The manipulation in this experiment involved assigning each stimulus-response pair to one of fourteen conditions. This assignment was made randomly for each pair, but was the same for all reinforcements of that pair. All conditions were run simultaneously; that is, the set of items that were active at any time included ones assigned to many different conditions. The fourteen conditions resulted from combinations of three independent variables affecting reinforcement: (1) The first of these variables was the delay itself. The presentation of the stimulus was terminated by the response, then the feedback (reinforcement) appeared, either immediately or following a delay of three, six, or twelve seconds. (2) During this delay, the subject was either allowed to do as he pleased or was instructed to count backwards from a randomly selected three-digit number. These conditions will be referred to as the *no-count* and the *count* conditions. (3) The feedback consisted either of the correct digit response presented alone or of both the stimulus trigram and the correct response. These conditions will be referred to as the *feedback-only* and the *stimulus-plus-feedback* conditions. In either case the duration of the reinforcement was four seconds. When the delay is zero, the count and no-count conditions are the same, hence only fourteen conditions are possible, instead of the $4 \times 2 \times 2 = 16$ conditions which might be expected.

The primary dependent variable considered in the experiment was the proportion of correct responses averaged over trials 2 through 7 (the initial trial, of course, was a random guess and has not been included in the average). In Fig. 4-10 this proportion is plotted as a function of the delay for the various reinforcement conditions. This figure shows all three of the trends which were mentioned above: the count, feedback-only condition shows a drop in the mean proportion correct as a function of delay;

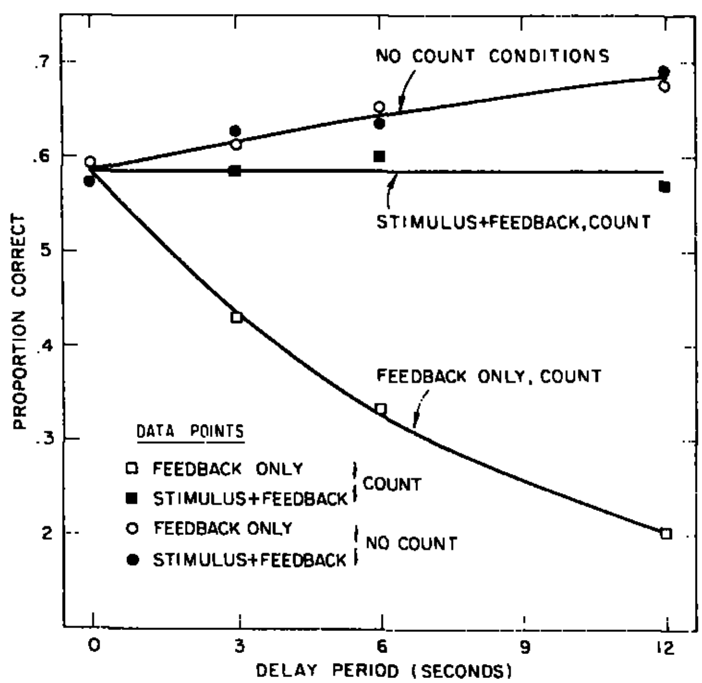


FIGURE 4-10. Observed and predicted probabilities of correct responses as a function of delay for two types of feedback and two types of delay activity.

the count, stimulus-plus-feedback condition shows no effect of delay; while both of the no-count conditions show an improvement with delay.

In interpreting the effects of reinforcement delay here, it is important to realize that the roles of rehearsal and of LTS are quite different in this task than they were in the two previous experiments. The presentation schedule was constructed so that there was always a substantial lag between successive appearances of an item. Because of this it was not practical for the subject to use a rehearsal buffer to maintain information until a response was required — too many of the items which intervened between study and test would have had to be ignored altogether. Instead, subjects were forced to rely primarily on LTS as a source of information storage. In such a case, subjects usually do not form a rehearsal buffer, but instead try to code each item as it is presented, and then turn their attention to the next item when it appears. The use of unique and relatively unfamiliar stimuli for each pair also increased the likelihood that this coding scheme was used.

The results of the count conditions are now fairly simple to interpret. The counting procedure had the effect of preventing rehearsal of information in STS; in particular, the subject could not readily remember the stimulus that was presented throughout the course of the delay period. Thus, in the feedback-only condition, the subject would frequently be unable to remember the stimulus by the time feedback was presented and would, therefore, be unable to associate the stimulus-response pair. In such a case, the probability of a correct response would drop toward chance as the likelihood increased that the stimulus could not be remembered; that is, as the delay interval increased. In the stimulus-plus-feedback condition, forgetting the stimulus during the delay period should have no effect since both members of the pair would always be available at the time of study. The counting task would, however, prevent any other processing from occurring during this interval, so the delay would be expected to have no effect at all.

In the no-count conditions the subject should have no problem in retaining the stimulus in STS during the delay interval; consequently, there should be no differences between the stimulus-plus-feedback and the feedback-only conditions. In fact, the delay interval can be spent in processing information in such a way as to make later LTS storage easier and more efficient. There are several ways in which this can be done; for example, the subject may engage in some sort of pre-processing of the stimulus, such as generating images or mnemonic codes which will aid in efficient storage once feedback is provided. Furthermore, after several reinforcements have been presented, the subject may be able to recover the response from LTS and recognize it as the correct one before the feedback is presented. He can then use the delay interval to further study the item. Either of these two processes can generate the increasing delay function that was observed.

Atkinson (1969) has described the amount of information which was transmitted to LTS by each reinforcement by an increasing exponential function for the no-count conditions and by a decreasing exponential function for the count conditions. These functions have been used to generate the predictions shown in Fig. 4-10. Although the sort of sequential investigations illustrated by Figs. 4-6 through 4-9 have not been made, the overall accuracy of these predictions support the interpretation.

The above analysis was able to accommodate effects that at first appeared to be inconsistent into a fairly simple framework by focusing attention on the informative value of the reinforcement, rather than treating it as a simple event. A similar, if not identical, analysis, we feel, will be able to reconcile the discrepant results that have been found for the effects of delay of reinforcement by other workers. It is experimental results of this sort that make a particularly strong case for our contention that factors involved in learning and memory are fundamental in determining the phenomena of reinforcement rather than the other way around.

Concept identification. In the following section, the theory will be applied to a concept-identification paradigm in which the effects of reinforcing events are quite different from those that have been discussed so far. The concept-identification task requires the subject to observe a series of stimuli and to classify them, one by one, into a fixed set of categories. Following each response, the subject is told the correct classification of the stimulus, and it is this feedback that gives rise to learning. The concept-identification procedure differs from the paired-associate procedure in that the classification depends systematically on some property (or properties) of the stimuli. This means that once the subject has solved the problem and has learned the rule by which stimuli are classified, he will be able to classify a novel stimulus correctly. There are, of course, an infinitely large number of possible stimulus properties and rules that can be used to partition the stimuli. In the experiment to be discussed below, we shall treat only a very few of these possibilities, those where the stimuli are composed of orthogonal binary dimensions and where the classification rule depends on only one of these dimensions. The procedure for the experiment that will be discussed (for a complete treatment see Wickens, 1969) will show these restrictions more clearly.

Subjects were seated before a teletype keyboard and saw stimuli projected on a screen in front of them. These stimuli were pictures which were constructed to vary along twelve different dimensions. Each of these dimensions, or attributes, of the pictures could take on either of two different values, only one value in each picture. One set of stimuli, for example, consisted of line drawings of houses in which the dimensions were represented by one or two windows, by a chimney on the left or on the right, and by ten other distinctions. From the twelve attributes a total of $2^{12} = 4{,}096$ distinct stimuli could be constructed. The rules used to determine the correct classifications were based on exactly one of these attributes; all stimuli for which that attribute took one value falling into one of two categories, all stimuli for which it took the other value falling into the other category. As each stimulus was presented, the subject indicated his choice of category by pressing the zero or the one key on the keyboard and was informed of the correct alternative by indicator lights mounted above the keyboard. A series of such trials was presented to the subject, the series continuing without interruption for the duration of a session. Whenever the subject had correctly identified the relevant attribute, as indicated by a string of twelve consecutive correct responses, he was signaled that the current problem was complete and was immediately started on a new problem, using a rule based on one of the eleven attributes that had not just been used. Subjects were run for two hours per day for five days. The number of problems solved by a subject during the experiment ranged from 53 to 115. During the first 25 problems or so, subjects showed improvement. After this point, however, the number of trials to solution remained approximately constant. The analysis to be discussed below is based on this stable, asymptotic data only.

The analysis that will be made of concept-identification is based upon the general idea of hypothesis testing (Bower and Trabasso, 1964; Restle, 1962). We assume that the subject solves concept problems by formulating hypotheses about the rule that determines the classification, then observing the sequence of classified stimuli to see whether the hypothesized rule is supported or not. A rule which is consistent with the true classification will enable the subject to respond correctly and thereby to solve the problem, whereas a rule that is inconsistent will cause errors to be made. When an inconsistency appears, the subject will abandon the rule under test and select a new one. It is apparent that this sort of solution is composed of two different processes: the selection of rules and their test. This dichotomy will represent an important part of our analysis of the role of reinforcement in concept identification.

We assume that initially there is a set of hypotheses which the subject considers to be potential solutions to the problem and which he wishes to test. The size of this pool depends on the nature of the task and on the subject's familiarity with it. In his first attempt to solve a concept-identification problem, a subject may have a large set of hypotheses which he views as possible, many of which hypotheses are quite complicated and cannot be the true solution to the problem. In the case of the experiment mentioned above, in which considerable practice was given and the subject was adapted to the task, the set of hypotheses may reasonably be identified with the set of attributes of the stimuli. In the following discussion, we shall speak of sampling *attributes*, indicating the specific nature of this experiment. One may, however, think of this as sampling from a pool of much more general hypotheses.

When solving a concept-identification problem, it is assumed that the subject starts by choosing a sample of r attributes from the total set and maintains them in STS by rehearsal. The matching of the values taken by these attributes to the two response alternatives is assumed to show local consistency (Gregg and Simon, 1967); that is, the assignment is made in such a way as to be consistent with the outcome of the last trial that has taken place. By comparing this assignment to the values that these attributes take in a new stimulus, the subject makes several predictions regarding the outcome of the new trial. Each of these predictions is based on one attribute in the sample: If the value of this attribute is the same as the value it took in the previous stimulus, then the same classification is predicted; if the value is different, then the classification is predicted to change. If more than two attributes are sampled, it is possible that the set of predictions may have internal inconsistencies, since each attribute may be varied independently of the others. The subject's classification response is generated from these predictions in some manner or other. The actual method of generation is not crucial to our analysis: He may choose a prediction at random, may select the response indicated by the largest number of predictions, or may use any of several other strategies.

The outcome of the trial provides confirmation of some of these predictions and disconfirmation of others, implying that those attributes on which incorrect predictions were based are no longer tenable candidates for the solution. Accordingly, these attributes are dropped from the rehearsal buffer. On the following trials, this process is continued, either until the buffer is emptied or until the problem is solved, in the sense that only one attribute is being considered and this is the correct one. If the buffer is emptied, the subject is forced to draw a new sample of attributes for testing. Here, for the first time, LTS becomes important. While the first set of attributes was being tested, information about them was being transferred to LTS. Now, when resampling is taking place, this information in LTS may allow the subject to avoid resampling those attributes which have already been tested and rejected. Resampling of an attribute that has already been tested may take place, but only when information about that attribute cannot be recovered from LTS, either because only a small amount of information was originally transferred or because of a failure of the search process. As more and more samples are drawn, there will be a greater and greater likelihood that the correct attribute will be selected and the problem solved.

The formulation of concept-identification learning given here is similar to a number of those that have been discussed in the literature, although it is not identical to any of them. In addition to the reference mentioned above, Trabasso and Bower have presented models in which questions of the delay of resampling (Trabasso and Bower, 1966) and the size of the test sample (Trabasso and Bower, 1968) have been discussed, while Gregg and Simon (1967) have considered a series of models which make a number of different assumptions about the selection of new hypotheses for test. All of these models, however, are different from our model in one critical respect, for they assume that the occurrence of an incorrect response causes the whole sample to be eliminated and redrawn. In contrast to this assumption, our theory makes a clear distinction between the effects of information feedback and the effects of reward. The important variable in determining what learning takes place is not whether the overt response was correct or in error, but rather the way in which the various predictions about the attributes were confirmed or disconfirmed. Since the subject can make a response that is not consistent with some of his predictions, it is possible for these predictions to be disconfirmed, and therefore rejected, at the same time that the response is correct. Only in the case where the buffer size is one (i.e., only a single attribute is under test) will the reward and information feedback aspects of the reinforcement be equivalent.

The fact that resampling does not take place on every error is central to our analysis of the role of reinforcement in this situation. It is relatively easy to demonstrate that this cannot occur as frequently as do errors. If resampling is postulated to take place after every error, the rate of learning for problems based on a particular attribute is independent of the value of r and can be represented by the probability that no more errors follow a given error; that is, by the probability that the correct attribute is both selected for rehearsal and is used as the basis for response generation. This solution probability can be estimated from the number of errors required to solve the problem. If m_i is the mean number of errors to solve problems based on the ith attribute,, then the solution probability for that attribute, c_i, can be estimated as follows (Restle, 1962):

$$\hat{c}_i = \frac{1}{m_i + 1}$$

The c_i's should form a probability distribution over the set of attributes. Using data from repeated problems for a typical subject, Wickens (1969) was able to determine $\hat{c}_i$ for all twelve attributes in the stimulus. These estimates summed to 1.8, which was significantly larger than the maximum value of 1.0 that would be permitted for a true probability distribution. The conclusion must be that the subject was learning more rapidly than could be accounted for by a process that depended only on whether the response was correct or not. Subjects must have used rehearsal buffers with sizes that were greater than one and must have depended on outcome information to adjust the contents of STS.

In his treatment of the data from this experiment, Wickens used a somewhat simplified version of the LTS postulate put forward in the preceding paragraphs; indeed, he did not separate his analysis into short- and long-term components as we have done. He assumed that all items contained in a particular sample were unavailable to the next ℓ samples, where $\ell = 0, 1, 2, \ldots$, and that this value of ℓ was constant for all attributes.[11] Using these assumptions, he was able to derive the distribution of the trial of last error and of the total number of errors, parametrized by combinations of r and ℓ. Figure 4-11 presents predictions for the mean trial of last error and compares them with the observed mean trial of last error for each of the forty-five subjects who served in the

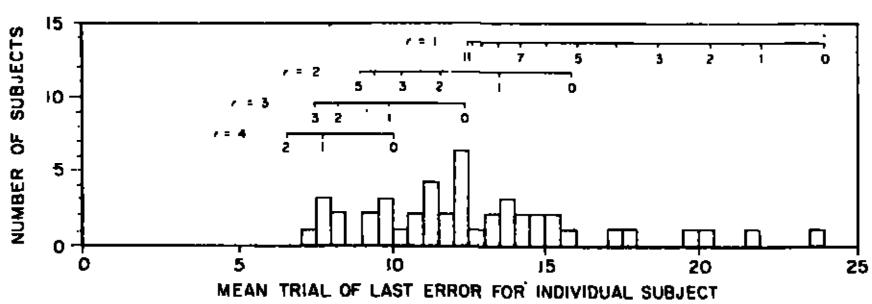


FIGURE 4-11. Frequency distribution of the mean trial of last error for individual subjects on a simple 12-dimensional concept-identification problem. Upper axes show theoretical predictions for four buffer sizes (r = 1, 2, 3, 4) and an appropriate range of delays in sampling replacement.

[11]The model that we have proposed above would predict that items from the same sample could remain unavailable for different lengths of time, and that these periods should depend upon the number of trials that the attributes resided in the rehearsal buffer.

experiment. The observed means are plotted as a histogram at the bottom of the figure, while the predictions are plotted along four short axes; a separate axis for $r = 1, 2, 3$, or 4. Points along these axes indicate values of ℓ. For example, there were three subjects whose mean trial of last error over all problems fell between 9.5 and 10.0. Mean trials of last error in this range are predicted by strategies in which $r = 4$ and $\ell = 0$, in which $r = 3$ and $\ell = 1$, or, to reasonable accuracy, in which $r = 2$ and $\ell = 4$. None of the strategies with $r = 1$ would be satisfactory for these subjects since, even with perfect long-term retention ($\ell = 11$), a mean trial of last error smaller than about 12 would be extremely unlikely. It is apparent from Fig. 4-11 that there is a very large spread in the observed data and that no single set of parameters can adequately account for all of the subjects. It is clear, however, that subjects with low values for the mean trial of last error were using strategies which required an r of at least 3 or 4 and which made significant use of LTS. The presence of these subjects who used rehearsal buffers of larger than a single attribute is again evidence for our contention that it is the confirmation of predictions about the attributes rather than the reward of a response that dictates the course of learning.

Magnitude of reward. The amount of reward associated with a correct response or the punishment associated with an error are variables that have not received a great deal of systematic consideration in human learning. In general, the studies that have examined amount of reinforcement have varied the degree of information feedback made available to the subject after his response (e.g., Keller, Cole, Burke, and Estes, 1965) or the amount of time that he is given to study the item (e.g., Keller, Thomson, Tweedy, and Atkinson, 1967). When reward magnitude has been considered, however, the extent of its effects seem to depend upon whether reward conditions have been compared between or within subjects. Several experiments by Harley (1965a,b) illustrate this clearly. He ran subjects in a paired-associate experiment using an anticipation procedure to learn CVC pairs. Incentive was provided for some pairs by telling the subject that he would receive twenty-five cents for each one that he correctly anticipated on a later trial. In one experiment (1965b), Harley tested for the effects of this reward in an absolute manner by comparing two groups of subjects: One group received twenty-five cents for every correct anticipation, whereas the other group received no rewards at all. The rate of learning for these two groups was virtually identical (see Fig. 4-12). When both reward values were used simultaneously with the same subject, half of the pairs receiving a reward and half not, the rewarded items were correct significantly more often (Harley, 1965a). As Fig. 4-12

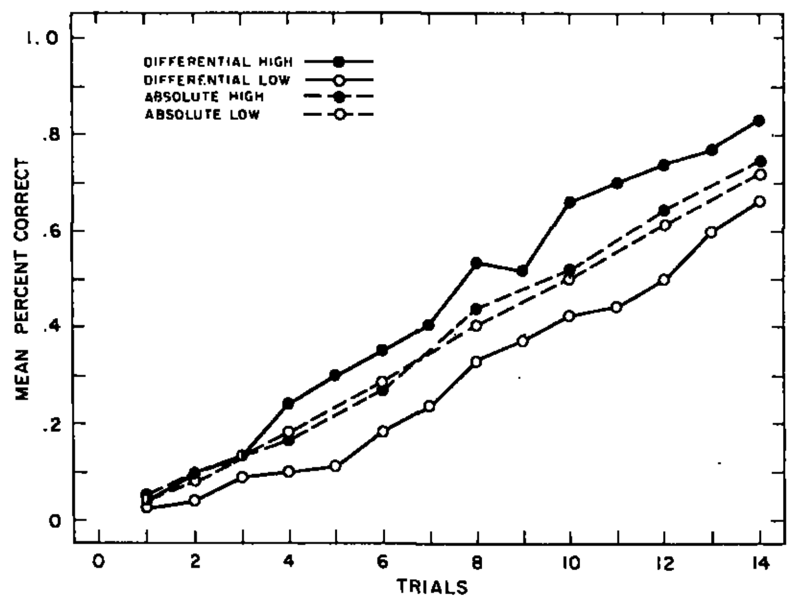


FIGURE 4-12. Learning curves for high- and low-rewarded paired-associate items tested with both reward values present at the same time differential procedure) or with values presented alone (absolute procedure). Data is replotted from 4-second groups in Harley 1965a, b).

indicates, this effect appears to take the form of an improvement in performance on the rewarded items and a decrement in performance on the unrewarded items when compared to either of the absolute groups. This interpretation is placed in some doubt by a later experiment (Harley, 1968) which suggests that the reward effect should be attributed primarily to poorer performance on the low-incentive items rather than to an improvement on the high-incentive items. In any case, these experiments indicate that the relative reward was the important variable, not the absolute magnitude of the reward.

In the system of reinforcement considered here, the reward associated with an item can influence performance only by altering the way in which information about the item is processed in STS. With this view, it is relatively easy to see why absolute rewards may not be important. The subject in a typical verbal-learning experiment is usually motivated to perform well, even in the absence of monetary incentive. The way in which information is processed in STS will be determined primarily by the nature of the test material and by the structure of the experiment. A difference in the absolute reward level will not make very much change in this scheme. When items with different reward values are presented, however, they may receive different treatments within the same general scheme. In particular, for tasks in which a rehearsal buffer is set up, the effects of differential rewards will be reflected in the relative probabilities of entering an item into the buffer or of deleting it once entered. Thus, high-reward items would be more likely to receive study than low-reward items, and so would be learned better. When only a single level of reinforcement is present, however, all items are equally likely to receive study, regardless of the level of reinforcement. The overall rate of learning in either case will be determined by the nature of the material to be learned and will not depend on the reward.

We have said that the effects of reward are determined by differences in the processing of high- and low-value items in STS. If this is the case, the nature of the reward effect should be influenced by the presence or absence of a rehearsal buffer. When a buffer is used, differential processing of high- and low-value items can occur easily, since high-point items may be entered into the buffer with a higher probability than low-point items, while low-point items (if recalled as such) may be more likely to be deleted from the buffer. On the other hand, if a coding strategy (similar to the one induced in the delay of reinforcement study) is used, each item will be studied as it is presented and there will be relatively little opportunity for an effect of reward magnitude to appear. Fortunately it is possible to predispose the subject to use either a rehearsal or a coding strategy by a fairly simple experimental manipulation. This effect has been demonstrated clearly in an experiment by Atkinson, Brelsford, and Shiffrin (1967) using two groups of subjects in a continuous paired-associate task in which number-letter pairs were given single reinforcements. In one group a fixed set of stimuli was used, pairing new responses with each stimulus throughout the course of a session. In the second group each stimulus was used only for a single pair, then retired (these two presentation procedures will be discussed more fully in the next paragraph). For the first group, clearly separate lag curves were obtained by varying the number of pairs that the subject was required to keep track of at any point in time; for the second group there was no effect of this manipulation on the lag curves. This difference is readily explained by assuming that subjects in the first group set up a rehearsal buffer, while subjects in the second group attempted to code each item during the interval before the presentation of the next pair.[12]

An experiment which looks at reward effects while manipulating the stimuli in this way has been conducted by Kirk Gibson at Stanford University. The paradigm of this experiment was, in general, similar to those that we have already analyzed. Subjects were seated at teletypes and were presented with a series of pairs to be learned. The stimuli were CVC trigrams and the responses were the letters of the alphabet. Each pair received only a single study and a single test. Two groups of subjects were run: In the *fixed-stimulus* condition a set of nine stimuli were selected at random at the start of each session and were used throughout that session. After each test in this condition, the same stimulus was presented for study paired with a new response. The second group of subjects was run in a *variable-stimulus* condition. In this condition, the item just tested was permanently discarded and a new stimulus-response pair was presented during the study phase of the trial. As in the fixed group, however, the subject was trying to keep track of only nine stimulus-response pairs at any given point in time. The same random presentation schedule employed in most of the other experiments was used, so that the test lags were distributed geometrically beginning with lag zero.

The second aspect of the experiment concerned the reward values assigned to the pairs. As each new item was presented for study, a value of either 11, 22, or 99 points was randomly assigned to it (i.e., each of these three values was equally likely to appear). The values were assigned independently for each item; in particular, a stimulus in the fixed group could receive different reward values when paired with different responses. The subject was told that if he correctly recalled an item, its points would be credited to his score for the session. At the time of test, the subject was not shown the point value associated with the item. Indeed, subjects were given no immediate feedback on their accumulation of points, although at the start of each session they were informed what percentage of the total possible points had been obtained during the previous session. The subjects were paid for participation in the experiment in proportion to this percentage.

The results of this experiment are shown in the form of lag curves in Figs. 4-13 and 4-14. For the fixed-stimulus group (Fig. 4-13) there was

[12]In their original paper Atkinson et al. (1967, p. 295) interpreted the difference in the two conditions by assuming that, for the second group, items were maintained in the buffer even after they had been tested. In light of later evidence, it now appears that this explanation is unrealistic and that the results may be more reasonably explained, as we have done, by the failure to form a buffer.

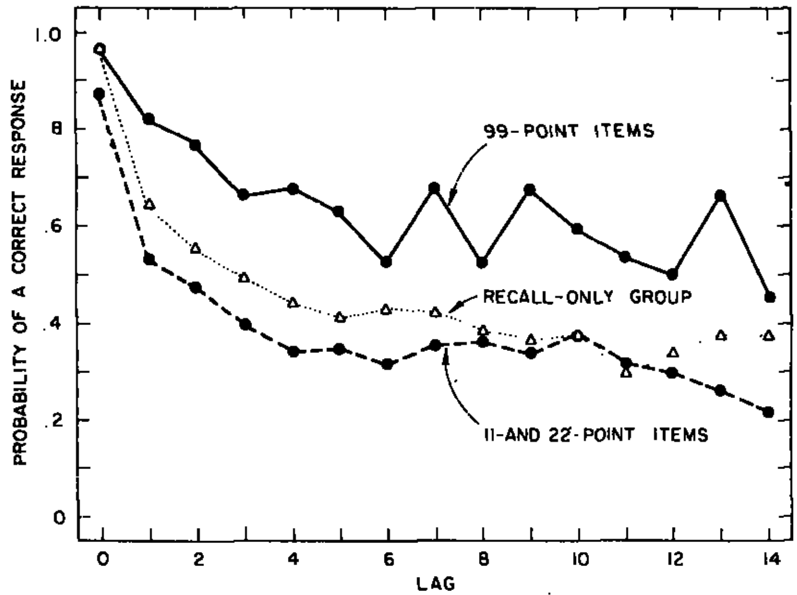


FIGURE 4-13. Probability of a correct response as a function of lag for items receiving different amounts of reward. The stimuli were a fixed set of trigrams.

a marked difference between performance on the 99-point items and on the other two types of items, although there was not a statistically significant difference between the 22- and the 11-point items. In contrast to these results there were no differences among the payoff conditions for the variable-stimulus procedure (Fig. 4-14). Apparently, varying the stimuli was sufficient to eliminate the basis for any reward effect.

The results of this experiment are in accord with our view of learning and reward. As indicated by subject reports at the conclusion of the experiment, the variable-stimulus pairs (a unique stimulus trigram and response letter) were fairly easy to code on an item-by-item basis. For this material, however, the subject experienced difficulty if he tried to maintain several items simultaneously in STS via rehearsal. Since it was much easier for the subject to code the items than to maintain a rehearsal buffer, he tended to study each item when it was presented and then turn his attention to the next item. Using this strategy, every item will be studied and the point

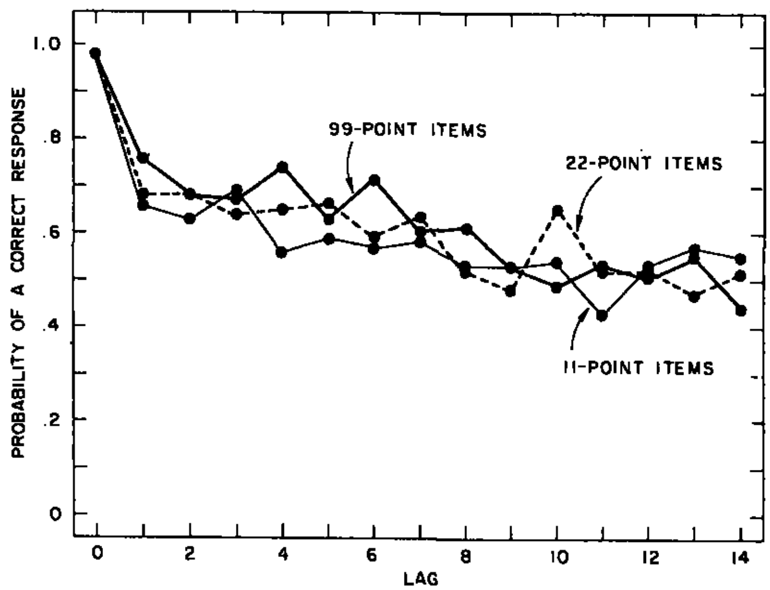


FIGURE 4-14. Probability of a correct response as a function of lag for items receiving different amounts of reward. A unique stimulus trigram was used for each item.

values will not play an important role in the amount of information transferred to LTS. Consequently, little or no effect of reward value should be observed, as indeed was the case for the variable-stimulus procedure.

On the other hand, for the fixed-stimulus procedure, the set of stimuli quickly became very familiar, and subjects reported that it was easy to set up a rehearsal buffer of three to five items. Coding, however, was much more difficult for this procedure, since it is almost impossible to generate noncompeting codes for the same trigram paired with many different letters during the course of a session. For this group, then, several items will be maintained in STS at any given time, and it will be easy to give preferential study to an item in the buffer by ignoring another item just presented. Similarly, a high-point item will almost always be entered into the buffer at the expense of some item that is already there. Thus the reward values will determine which items are studied and for how long they are maintained. Accordingly, a reward effect is predicted for the fixed-stimulus procedure, as was observed.

We do not want to argue from these results that a reinforcement effect cannot be obtained using the variable-stimulus procedure. If sufficiently large rewards are offered for correct responses to certain items, then there is no doubt that they will receive additional study, probably both by rehearsal and by coding. The point that we feel is important here is that with the particular payoff levels used in the study, a marked difference in reinforcing effects appeared between the fixed- and variable-stimulus procedures, two procedures which in a logical sense place identical demands on a subject. Although both procedures require the subject to keep track of the same number of stimulus-response pairs at any given point in time, the particular nature of the stimulus material caused different methods of study to be used, and in turn made reinforcement effects evident in one case and not in the other. This is another example where a given reinforcing operation can lead to markedly different effects depending on the particular information-processing requirements of the learning task.

One interesting feature of the experiment is the high accuracy of recall obtained for the variable-stimulus condition. Although there was no effect of the reward, the overall proportion of correct responses is approximately at the same level as the 99-point items for the fixed-stimulus group. This presumably reflects the fact that stimulus-response pairs in the variable-stimulus condition are less subject to interference from other pairs than in the fixed-stimulus condition. Further studies are currently in progress to investigate the exact form of the STS structure that is set up for the two conditions.

It is not possible to make a direct comparison of rewarded and unrewarded performance within this study. Some sort of comparison can be made, however, between another of Gibson's groups and a group from the experiment by Loftus reported in the first part of this paper. The group in question used a fixed-stimulus procedure, but with the digits 1 through 9 as stimuli instead of trigrams. This procedure is exactly the same as the recall-alone condition of the Loftus study, except for the presence of rewards. If these rewards are neglected, performance in the two experiments is almost exactly the same; if the three reward values are combined, the mean lag curve is indistinguishable from that observed by Loftus. The unrewarded responses of the recall-alone condition fall roughly between the items which had been given high and low incentives (see Fig. 4-15). In this figure the 11- and the 22-point items have been combined, hence each data point in this curve includes approximately twice the number of observations as the corresponding point in the high-reward curve (this means that the average of the two curves does not lie midway between them; in fact it falls almost exactly on the curve for the recall-alone group). While hardly conclusive, this comparison again suggests that the

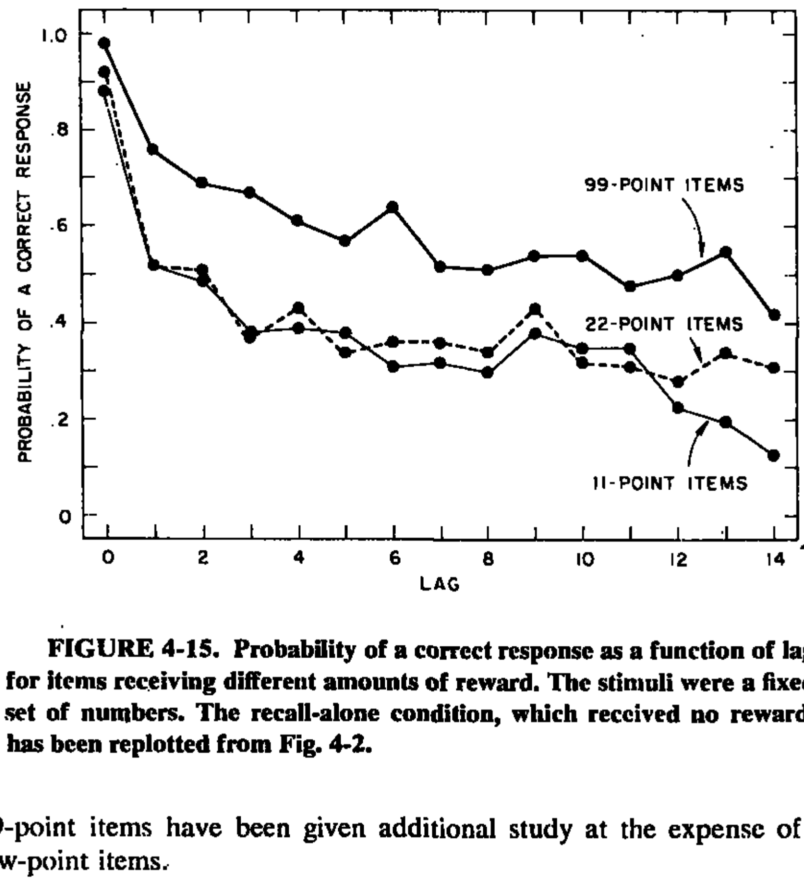


FIGURE 4-15. Probability of a correct response as a function of lag for items receiving different amounts of reward. The stimuli were a fixed set of numbers. The recall-alone condition, which received no reward, has been replotted from Fig. 4-2.

99-point items have been given additional study at the expense of the low-point items.

Effects of reinforcement on retrieval. Throughout this paper, a distinction has been made between storage and retrieval processes in learning. As noted in the introduction, this distinction is also relevant to an analysis of reinforcement. The applications considered so far have been primarily concerned with how reinforcement influences the study of items, hence the storage of information. The reason for not turning sooner to retrieval aspects of reinforcement is that there are few experiments dealing specifically with this topic (Wasserman, Weiner, and Houston, 1968; Weiner, 1966).

In an attempt to remedy this state of affairs, we have initiated some experiments in which the reward associated with paired-associates has been manipulated both at the time the item is first studied and later at test.

None of these experiments is yet complete, but we want to present some pilot data from an experiment by Geoffrey Loftus which illustrate some effects of interest. This experiment employed a continuous memory task that was almost identical to the fixed-stimulus procedure described in the section on reward magnitude. The stimuli were the digits from 1 to 9, and the responses were letters of the alphabet. Each new stimulus-response item was assigned a value of either 11, 22, or 99 points. When an item was presented for study, however, its point value was not always displayed. For about half of the items, no information about the reward was given at this time; the subject was instructed that the items for which no point values appeared had, nevertheless, been assigned one of the three values at random by the computer controlling the experiment, and that these values would count in his total score for the session. Similarly, when the items were tested, their reward value might or might not be displayed. Again, the reward value was presented on about half of the tests. The presentation of the reward value at test was independent of whether the reward had been presented during study; thus the subjects might receive information about the rewards assigned to a particular item at the time of study, at the time of test, at both times, or at neither time. If a reward value was presented at study and test, then the same value appeared both times.

Some preliminary results from this study are presented in Fig. 4-16. The graph gives the proportion of items correctly recalled, averaged over all test lags, as a function of the presentation schedule and reward value. The mean latencies of correct and error responses are also shown. As in Gibson's experiment, there was very little difference between the 11- and 22-point items, so these have been grouped together as low-value items. The two points on the left of the graph are for the conditions in which the subject was informed during study that he was being shown a high (i.e., 99) point item. One of the observations (HH) shows the results when the reward information was also presented at test, the other (H-) when it was not. Similarly, the three middle points (-H, – –, -L) are associated with conditions in which no reward was presented at the time of study, while the two right-most points (L-, LL) give results for items studied with a low-point value (11 or 22). Although all test lags have been combined in this figure, the general form of the results appears to be the same at both short and long lags.

The major effects in Fig. 16 are due to the reward values displayed during study. Items that were assigned 99 points at study had a higher probability of being recalled than items for which no reward value was assigned. These items were, in turn, better remembered than the low-point items. The explanation that we offered for Gibson's data in the previous section is consistent with these findings if items with an unspecified reward

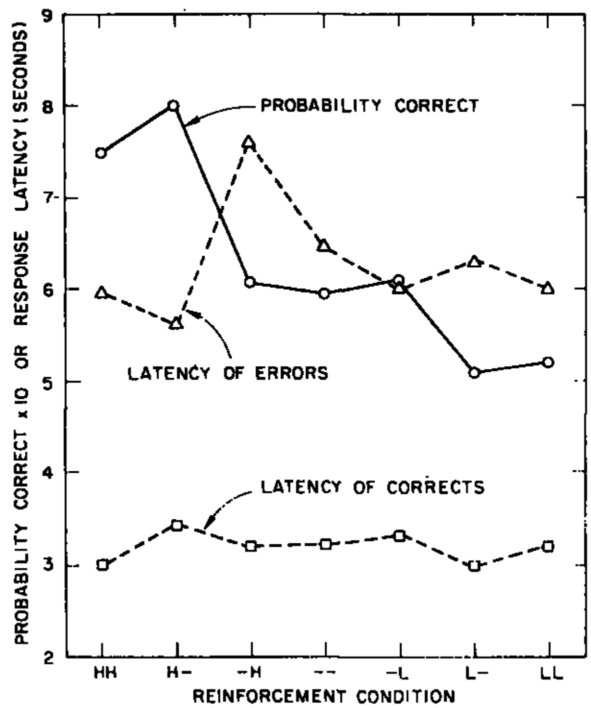


FIGURE 4-16. Probability of a correct response and latency of correct and error responses as a function of reward information given at study and test. The first letter in the condition label designates reward at study, the second designates reward at test; H indicates 99-point reward, L indicates 11- or 22-point reward, — indicates that no reward information was given.

are assumed to receive a level of study intermediate between that given to high- and low-point items.

In the introduction, two ways were mentioned by which reinforcement could aid retrieval. The first of these suggested that the reward value associated with an item might act as a cue to facilitate the retrieval of information from LTS. These preliminary data provide little support for this hypothesis, for there is no indication that items for which the reward value was presented on both study and test are better recovered than those that received reward only at the time of study. This result indicates that in this experiment the reward had negligible cue value. The second potential effect of reward on retrieval receives more support; namely, that a subject would be willing to spend more time in attempting to retrieve items that had been assigned a high value than items that had been assigned low values. This effect is quite clearly shown in the latency of incorrect responses, particularly for the conditions in which the reward value had not been identified during study (i.e., conditions -H, – –, and -L). The latency of errors shows the same effect for the two conditions where point values were presented during study, although not to as marked an extent. Curiously, this effect is totally lacking in both the latency and probability of a correct response. These results suggest that either the subject was able to retrieve an item without much difficulty (with a latency of about three seconds), or else no recovery was possible. When an item could not be recovered, the additional search time spent on items with large reward values was not of much help. There was no limit on the time that was available to make a response, so the failure to retrieve cannot be attributed to a premature termination of the trial.

These results must be regarded with some caution. The amount of data represented is not great, and it is likely that the specific characteristics of the task are not optimum for demonstrating retrieval effects. The fixed-number procedure that was used is one which almost invariably leads the subject to set up a rehearsal buffer. Indeed, several of the subjects reported being able to successfully set up a nine-item buffer by visualizing the responses arrayed in a 3 × 3 matrix! The process of retrieving items from the buffer is a fairly simple one and invariably will lead to a correct response. Items that are recovered in this manner will not contribute to any effects of reinforcement on the recovery of the item. We would expect that more substantial effects will be observed in a task in which the subject is forced to put greater reliance on LTS. Nevertheless, an effect of reinforcement on retrieval time was clearly evident in this study, showing, as expected, an incentive effect. This effect would not be predicted from a theory that assigned to reinforcement only the role of strengthening connections; it is, however, consistent with the view that reinforcement acts to direct attention and to control information flow.[15]

Conclusion

In this paper we have attempted to present a theoretical framework within which to view the phenomena of reinforcement. Basically, the

[15]A replication of this experiment (Loftus and Wickens, 1970), using a slightly modified procedure, demonstrated effects of study and test cueing of incentive on both the probability of a correct response and on response latency. These results are in agreement with the analysis presented here.

framework involves an account of learning and attention in terms of the storage of information in memory and its subsequent retrieval. Reinforcement is the modulation of this information flow as it influences both storage and retrieval processes. It is our belief that a given reinforcing operation can have many different and often seemingly contradictory effects depending on the particular study and test procedures that are used. In order to illustrate some of these effects, the theory was applied to results from several different experimental paradigms. These applications, we hope, have demonstrated the general principles by which the transfer of information in memory is controlled and shaped by reinforcement.

It is unfortunate that our discussion of reinforcement cannot be summed up in the form of a set of simple statements. Statements of this type, such as that of the law of effect, do not provide a consistent and unambiguous explanation of the range of reinforcement phenomena that have been observed. If the effects of reinforcement are analyzed in the context of an information-processing theory of the type outlined in this paper, we believe that they will appear relatively orderly and consistent.

REFERENCES

Atkinson, R. C. Information delay in human learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 507–511.

Atkinson, R. C., Brelsford, J. W., Jr., and Shiffrin, R. M. Multiprocess models for memory with applications to a continuous presentation task. *Journal of Mathematical Psychology*, 1967, *4*, 277–300.

Atkinson, R. C., and Shiffrin, R. M. Mathematical models for memory and learning. Technical Report 79, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California, 1965. (To be published in D. P. Kimble (Ed.), *Proceedings of the third conference on learning, remembering and forgetting.* New York: New York Academy of Sciences.)

Atkinson, R. C., and Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. 2. New York: Academic Press, 1968. Pp. 89–195. (a)

Atkinson, R. C., and Shiffrin, R. M. Some speculations on storage and retrieval processes in long-term memory. Technical Report 127, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California, 1968. (b)

Bilodeau, E. A., and Ryan, F. J. A test for interaction of delay of knowledge of results and two types of interpolated activity. *Journal of Experimental Psychology*, 1960, *59*, 414–419.

Bourne, L. E. Information feedback: Comments on Professor I. McD. Bilodeau's paper. In E. A. Bilodeau (Ed.), *Acquisition of skill*. New York: Academic Press, 1966. Pp. 297–313.

Bower, G. H. An association model for response and training variables in paired-associate learning. *Psychological Review*, 1962, *69*, 34–53.

Bower, G. H., and Trabasso, T. R. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford, Calif.: Stanford University Press, 1964. Pp. 32–94.

Brelsford, J. W., Jr., and Atkinson, R. C. Short-term memory as a function of rehearsal procedures. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 730–736.

Brelsford, J. W., Jr., Shiffrin, R. M., and Atkinson, R. C. Multiple reinforcement effects in short-term memory. *British Journal of Mathematical and Statistical Psychology*, 1968, *21*, 1–19.

Broadbent, D. E. Flow of information within the organism. *Journal of Verbal Learning and Verbal Behavior*, 1963, *4*, 34–39.

Buchwald, A. M. Effects of immediate vs. delayed outcomes in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 317–320.

Buchwald, A. M. Effects of "right" and "wrong" on subsequent behavior: A new interpretation. *Psychological Review*, 1969, *76*, 132–143.

Crowder, R. G., and Morton, J. Precategorical acoustic storage (PAS). *Perception and Psychophysics*, 1969, *5*, 365–373.

Elliot, P. B. Tables of d'. In Swets, J. A. (Ed.), *Signal detection and recognition in human observers: Contemporary readings*. New York: Wiley, 1964. Pp. 651–684.

Estes, W. K. Reinforcement in human learning. In J. Tapp (Ed.), *Reinforcement and behavior*. New York: Academic Press, 1969. Pp. 63–94.

Feigenbaum, E. A. The simulation of verbal learning behavior. In E. A. Feigenbaum and J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill, 1963. Pp. 297–309.

Festinger, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row Peterson, 1957.

Freund, R. D., Loftus, G. R., and Atkinson, R. C. Applications of multiprocess models for memory to continuous recognition tasks. *Journal of Mathematical Psychology*, 1969, *6*, 576–594.

Greenspoon, J., and Foreman, S. Effect of delay of knowledge of results on learning a motor task. *Journal of Experimental Psychology*, 1956, *51*, 226–228.

Gregg, L. W., and Simon, H. A. Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology*, 1967, *4*, 246–276.

Guthrie, E. R. Association of contiguity. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959. Pp. 158–195.

Harley, W. F., Jr. The effect of monetary incentive in paired-associate learning using a differential method. *Psychonomic Science*, 1965, *2*, 377–378. (a)

Harley, W. F., Jr. The effect of monetary incentive in paired-associate learning using an absolute method. *Psychonomic Science*, 1965, *3*, 141–142. (b)

Harley, W. F., Jr. Delay of incentive cues in paired-associate learning and its effect on organizing responses. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 924–929.

Hill, J. W., and Bliss, J. C. Modeling a tactile sensory register. *Perception and Psychophysics*, 1968, *4*, 91–101.

Hintzman, D. L. Explorations with a discrimination net model for paired-associate learning. *Journal of Mathematical Psychology*, 1968, *5*, 123–162.

Hochman, C. H., and Lipsitt, L. P. Delay-of-reward gradients in discrimination learning with children for two levels of difficulty. *Journal of Comparative and Physiological Psychology*, 1961, *54*, 24–27.

Kamin, L. J. Predictability, surprise, attention and conditioning. In B. A. Campbell and R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts, 1969. Pp. 279–296.

Keller, L., Cole, M., Burke, C. J., and Estes, W. K. Reward and information values of trial outcomes in paired-associate learning. *Psychological Monographs*, 1965, *79* (Whole No. 605).

Keller, L., Thomson, W. J., Tweedy, J. R., and Atkinson, R. C. Effects of reinforcement intervals in paired-associate learning. *Journal of Experimental Psychology*, 1967, *73*, 268–277.

Kintsch, W., and McCoy, D. F. Delay of information feedback in paired-associate learning. *Journal of Experimental Psychology*, 1964, *68*, 372–375.

Lawrence, D. H., and Festinger, L. *Deterrents and reinforcement: The psychology of insufficient reward*. Stanford, California: Stanford University Press, 1962.

Loftus, G. R., and Wickens, T. D. The effect of incentive on storage and retrieval processes. *Journal of Experimental Psychology*, 1970, *85*, 141–147.

Millward, R. B. An all-or-none model for noncorrection routines with elimination of incorrect responses. *Journal of Mathematical Psychology*, 1964, *1*, 392–404.

Montague, W. E., Adams, J. A., and Kiess, H. O. Forgetting and natural language mediation. *Journal of Experimental Psychology*, 1966, *72*, 829–833.

Murdock, B. B., Jr. Recent developments in short-term memory. *British Journal of Psychology*, 1967, *58*, 421–433.

Norman, D. A. Toward a theory of memory and attention. *Psychological Review*, 1968, *75*, 522–536.

Peterson, L. R., and Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, *58*, 193–198.

Phillips, J. L., Shiffrin, R. M., and Atkinson, R. C. The effects of list length on short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 303–311.

Postman, L. Rewards and punishments in human learning. In L. Postman (Ed.), *Psychology in the making*. New York: Knopf, 1962. Pp. 331–401.

Rescorla, R. A. Conditioned inhibition of fear. In W. K. Honig and N. J. Mackintosh (Eds.), *Fundamental issues in associative learning*. Halifax: Dalhousie University Press, 1969. Pp. 65–89.

Restle, F. The selection of strategies in cue learning. *Psychological Review*, 1962, *69*, 329–343.

Rundus, D. Paired-associate recall following a distractor task: Initial and second choice performance. *Journal of Mathematical Psychology*, 1970, *7*, 362–370.

Rundus, D., and Atkinson, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 99–105.

Saltzman, I. J. Delay of reward and human verbal learning. *Journal of Experimental Psychology*, 1951, *41*, 437–439.

Shiffrin, R. M. Search and retrieval processes in long-term memory. Technical Report 137, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California, 1968.

Shiffrin, R. M., and Atkinson, R. C. Storage and retrieval processes in long-term memory. *Psychological Review*, 1969, *76*, 179–193.

Simon, H. A., and Newell, A. Information processing in computer and man. *American Scientist*, 1964, *52*, 281–300.

Sperling, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, *74* (Whole No. 498).

Spring, C. Decay and interference theories of short-term forgetting. *Psychonomic Science*, 1968, *12*, 373–374.

Sternberg, S. High-speed scanning in human memory. *Science*, 1966, *153*, 652–654.

Thomson, W. J. Recall of paired-associate items as a function of interpolated pairs of different types. *Psychonomic Science*, 1967, *9*, 629–630.

Thorndike, E. L. *Human learning*. New York: Appleton-Century-Crofts, 1931.

Tolman, E. C. *Purposive behavior in animals and men*. New York: Appleton-Century-Crofts, 1932.

Tolman, E. C., and Brunswik, E. The organism and the causal texture of the environment. *Psychological Review*, 1935, *42*, 43–77.

Trabasso, T. R., and Bower, G. H. Presolution shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. *Journal of Mathematical Psychology*, 1966, *3*, 163–173.

Trabasso, T. R., and Bower, G. H. *Attention in learning: Theory and research*. New York: Wiley, 1968.

Wasserman, E. A., Weiner, B., and Houston, J. P. Another failure for motivation to enhance trace retrieval. *Psychological Reports*, 1968, *22*, 1007–1008.

Weiner, B. Motivation and memory. *Psychological Monographs*, 1966, *80* (Whole No. 626).

Wickens, T. D. Attribute elimination strategies for concept identification with practiced subjects. Technical Report 3, Laboratory of Human Learning, Brown University, Providence, R. I., 1969.