Spring 3-5-1955

An Experimental Investigation of Lashley and Wade's "The Pavlovian Theory of Generalization"

Allen B. Madans

"The Pavlovian Theory of Generalization" — Madaus

AN EXPERIMENTAL INVESTIGATION OF LASHLEY AND WADE'S "THE PAVLOVIAN THEORY OF GENERALIZATION"

by

Allen B. Madans

A Thesis

In partial fulfillment of the

Requirements for the Degree of

Master of Arts in Psychology

The University of New Mexico
1954

This thesis, directed and approved by the candidate's committee, has been accepted by the Graduate Committee of the University of New Mexico in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

E. A. Castetter
DEAN

3/5/1955
DATE

Thesis committee

Gordon Peterson
CHAIRMAN

Roger J. Weldon

Arthur J. Bahm

ACKNOWLEDGMENT

The writer wishes to express his thanks for the guidance, assistance and encouragement received from Dr. George M. Peterson and Dr. Roger J. Weldon while the writer was engaged in the investigation here reported. Further thanks are due to Robin O'Connell Madans for taking on a large part of the work entailed in the collection of data.

TABLE OF CONTENTS

CHAPTER I

THE PROBLEM DEFINED, ITS PAST AND PRESENT

I. A DEFINITION OF STIMULUS GENERALIZATION

In the following paper, the term "stimulus generalization" will appear many times. As a matter of convenience it will be abbreviated to "SG."

The best type of definition to give for SG is an operational one. Although the operational definition loses in generality, it gains by providing at least one concrete illustration of the phenomenon with which this paper deals. The following illustration is taken from the experimental literature in most part, and it is the type of illustration that is recognized both by those for and those against SG as demonstrating the phenomenon of SG.

Assume a Pavlovian dog preparation, in which a dog's salivation to the conditioned stimulus can be collected and accurately measured. Now suppose this dog is conditioned to salivate to a 1,000-cycle tone. It will be found that a dog so trained will also salivate to a 900-cycle tone, or to a 1,100-cycle tone, but not quite so much as it would to the original 1,000-cycle tone. And, further, the dog will respond by salivating to a 750-cycle tone, but not quite so much as it did to the 900-cycle tone, and not nearly so much as it did to the original 1,000-cycle tone. Frequency has its correlate in the psychological

dimension of pitch (by human standards); and when an animal is trained to respond to a specific stimulus, some specific point on a given dimension, the animal will also respond at least partly to stimuli on the same dimension that do not perfectly match the original stimulus; and the strength of this response will bear some lawful relationship to the physically determined distance of these non-matching stimuli from the original stimulus. This is what we mean when we talk about SG. The definition is admittedly oversimplified, but it is hoped that some of this deficiency will be made up in the course of the discussion.

Definitions of other terms will be given as they arise in the exposition of the argument.

II. A SHORT HISTORY OF SG

About 1910, Pavlov described a phenomenon which has come to be known as stimulus generalization.[1] Pavlov's initial observations of this phenomenon were not systematic, and he never personally investigated it further. Instead, Anrep (a student of Pavlov's) conducted a major study on SG in 1920 and 1923, using dogs as subjects.[2] (This is the same Anrep who was later to translate into English Pavlov's major work, Conditioned Reflexes.) On the basis of his study, Anrep concluded that SG was a genuine phenomenon.

Anrep's work stood practically unchallenged for close to 15

---

[1]Osgood, C.E. Method and theory in experimental psychology. New York: Oxford University Press, 1953, p.354.

[2]Ibid.

years. In 1933, Loucks[3] published a criticism of Pavlovian reflexology in general, and in it Anrep's study was called to question. Loucks doubted the authenticity of Anrep's data, and said as much. A year later, in 1934, "...being somewhat disturbed that such a presumably sound and important principle [SG] should be called in question even indirectly..." Bass and Hull set out to substantiate Anrep experimentally.[4] Bass' and Hull's results were quite positively in favor of Anrep. Following Bass and Hull, Hovland[5] in 1937 conducted a series of experiments, all of which lent support to the hypothesis of SG. More recently Wickens in 1943 and Spence in 1945 have offered additional evidence that SG is a genuine phenomenon.[6]

---

[3]Loucks, R.B. An appraisal of Pavlov's systematization of behavior from the experimental standpoint. J. comp. Psychol., 1933, 15, 1-45.

[4]Bass, M.J., & Hull, C.L. The irradiation of a tactile conditioned reflex in man. J. comp. Psychol., 1934, 17, 47-65.

[5]Hovland, C.I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. J. gen. Psychol., 1937, 17, 125-148.

------- The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. J. genet. Psychol., 1937, 51, 270-291.

------- The generalization of conditioned responses: III. Extinction, spontaneous recovery, and disinhibition of conditioned and of generalized responses. J. exp. Psychol., 1937, 21, 47-62.

[6]Wickens, D.D. Studies of response generalization in conditioning. I. Stimulus generalization during response generalization. J. exp. Psychol., 1943, 33, 221-227.

[7]Spence, K.W. An experimental investigation of Lashley's interpretation of primary stimulus generalization (in manuscript.)

Despite the great amount of evidence in its favor, SG is still not universally accepted. Loucks' 1933 article was the first serious argument levelled against SG, but Loucks' criticism was not based on experiment. The first serious, experimentally founded criticism -- to judge from the hornet's nest of neo-Pavlovian indignation that it stirred up -- came from Lashley and Wade in their 1946 article, "The Pavlovian Theory of Generalization."[7] The seriousness of this criticism stems not from any striking experimental refutation of SG; to the contrary, the experiment reported by Lashley and Wade (referred to hereafter as "L & W") to support their criticism does more to harm their argument than to help it. The criticism's claim to attention is based on the fact that (1) it comes, in part, from Lashley, a man whose outstanding work in other areas of experimental psychology entitles him to a hearing; (2) the article contains a number of arguments unfavorable to SG, and these arguments are cogent and worthy of note and answer by those for whom SG is an accepted phenomenon.

The following section will examine the main points of the L & W argument.

---

[7]Lashley, K.S., & Wade, M. The Pavlovian theory of generalization. Psychol. Rev., 1946, 53, 72-87.

III. LASHLEY AND WADE'S ARGUMENT: THE CIRCULARITY OF SG

In the opening paragraphs of their article, L & W refer to SG as "...a physically and physiologically meaningless spread of the effects of training along 'dimensions' of similarity of stimuli."[8] Note how carefully L & W set apart the word "dimension," by placing it in quotes. For the essence of the L & W argument lies in their statement that, at the beginning of the naive animal's training, there is no dimension at all as far as the animal is concerned.

Reminiscent in its insurrectionist spirit of the James-Lange theory of emotion, this unquestionably striking idea exemplifies a commendable propensity to challenge even the most reasonable-sounding idea. For is it not reasonable to assume that varying brightnesses are differentiated from varying loudnesses in like manner by both animals and humans? Or that varying frequencies form a psychological continuum for rats as well as for people? Is there no common denominator within classes of stimuli, for animals, which will bind the stimuli in one class (from a human standpoint) into what could be called a dimension of similarity? To the previous questions, L & W answer an emphatic "No," and they provide what they consider to be experimental evidence to support their contention.

L & W state that, to the naive animal, there is only one dimension of similarity. It is a difficult dimension to conceive of, since all stimuli in the animal's environment belong to it. Perhaps the

---

[8]Lashley and Wade, op. cit., p.73.

best explanation of this dimension can be given by L & W's description of what happens to the organism during classical conditioning.

> For the animal something happens [the conditioned stimulus] and then comes an electric shock. In the early stages of conditioning any change in the environment may elicit the avoiding reaction. Even with human subjects, conditioned to the sound of a bell, [Lashley] has obtained the conditioned reaction without further training from the sound of a buzzer, of breaking glass, of clapping hands, from a flash of light, from pressure or prick on arm or face.... The only 'dimension' common to such stimuli is that all produce a sudden change in the environment. [Italics not in the original.][9]

And further,

> ...the dimension itself is created by or is a function of the organism and only secondarily, if at all, a property of the physically definable character of the stimuli.[10]

From these points of view, it is only a short step to bringing the charge of circularity against SG theorists; L & W do not miss this opportunity. They first quote a statement from Hull's "Principles of Behavior":

> It is held that the number and nature of the various primary generalization gradients are caused jointly by the nature of the stimulus energy and the nature of the receptor response.[11]

L & W comment on this idea as follows:

> No theory is proposed, however, to account for the production of a 'gradient of afferent generalization' by the receptor response.

---

[9] Lashley and Wade, op. cit., p.80.

[10] Ibid., p.82.

[11] Hull, C.L. Principles of behavior. New York: D. Appleton-Century, 1943, p.198.

> The statement [quoted] therefore merely constitutes an assumption that, through some mysterious process, the effects of training irradiate to stimuli which are similar for the organism. Similarity is not defined other than as a product of this irradiation.[12]

Thus, the entire idea of similarity of stimuli, upon which rests the structure of SG theory, is "completely circular." By asserting the reality to the organism of similarity of stimuli, SG theorists are merely begging the question, state L & W.

Here, then, is L & W's argument in summary: It is claimed that animals trained to react to a single stimulus react to stimuli that are similar to the original stimulus, but that the amplitude of the reaction is a function of the distance of the second stimulus from the original stimulus. But in what way can we be sure that the stimuli are truly similar, so far as the animal is concerned? We know the stimuli are similar (for the animal) because he reacts in the differential fashion that we call stimulus generalization. Thus the tautology is complete, and the nature of SG becomes entirely dependent upon an indefinable phrase: "dimension of similarity."

The essence, then, of L & W's argument is that SG has no meaning apart from a tautologous one. Animals are said to respond to stimuli along a generalization gradient, if the stimuli which are being used to demonstrate this gradient are "similar." However, L & W claim, no adequate definition of similarity is offered by SG theorists, except by saying that if animals react to stimuli in this regular, graded fashion,

---

[12] L & W, op. cit., p.83.

then the stimuli are similar, and constitute a dimension of similarity for the given animal. Suppose, for instance, one of Pavlov's dogs, prepared for saliva measurements, is stimulated by the original tone to which he was trained, say a 1,000-cycle tone. To this 1,000-cycle tone he yields 1cc. of saliva. Now suppose instead of presenting another tone, a bright light is presented to the dog, and then suppose that the dog salivates to the extent of .9cc. of saliva, in response to the bright light. The animal has generalized its original response (to a tone) to a new stimulus, a bright light. Are we then to say, ask L & W, that sound and light are on a "dimension of similarity"? Since the existence of similarity is being judged by the behavior of an organism, we cannot escape drawing the conclusion that sound and light are somehow similar. SG theorists have been making a serious error, according to L & W, in classifying stimuli as similar by human standards, and then presenting these classes of stimuli to animals as if they must then be similar for animals. If SG investigators -- those who found positive evidence for the existence of an SG gradient -- had presented different types of stimuli (as judged by human standards) during their experimental investigations, they would have found dimensions of similarity far different from any they ever dreamed existed. The fault, it seems, lies in excess anthropomorphism; in this case, imputing to animals the ability or proclivity to perceive classes of stimuli in the same manner that humans perceive these stimuli. There is an implicit assumption in this procedure that a group of stimuli are similar for animals simply because they are similar for humans.

The force of the L & W argument is not inconsiderable, but a weakness appears in their position when L & W make a grudging admittance of the slim possibility of something that might just be confused with SG. L & W say:

> When the physical dimension is quantitatively continuous, as the frequency of sound or light waves, confusions of discrimination may be more frequent for remote than for adjacent points on the continuum...[13]

These "confusions of discrimination" sound suspiciously like the examples of the very phenomenon that SG theorists describe, and which they (the SG theorists) call SG.

Why do these "confused" responses occur with less amplitude the farther away they are from the original stimulus? L & W do not completely ignore the finding, by many experimenters, of a fairly regular SG gradient, but their explanation of this highly important experimental fact constitutes the weakest part of their argument. The weakness of the argument comes at a most inopportune point, because the existence of a gradient is one of the strongest pieces of evidence in favor of the SG hypothesis.

> The unsophisticated subject reacts explosively to any change in the environment; the more sophisticated subject may limit reactions to stimuli which past experience has shown to be significant within the experimental situation, but still without close attention. Consequently a test for irradiation may give the appearance of a gradient of habit strength when it is actually measuring discriminative thresholds under distraction.... In those cases where positive evidence for a gradient has been obtained, experimental conditions were such as to make discrimination difficult. When inattention and threshold values are not involved, no evidence of a gradient is found. [Italics not in the original.][14]

---

[13] L & W, op. cit., p.83.

[14] Ibid., p.84.

Thus, L & W are willing to rest their case on the matter of attention. To support their argument that inattention and threshold values are the determining factors in experiments that report positively on SG, L & W designed a study that would eliminate these factors. The apparatus employed was the Lashley jumping stand. The use of this apparatus eliminates the two troublesome factors in one fell swoop, since attention is forced, and discriminative thresholds have already been measured with some accuracy by Lashley.[15]

The last idea to consider under the heading of circularity, before going into a description of the L & W experiment and its repercussions, is the nature of the charge of circularity brought by L & W against SG theorists.

The pragmatic tradition of science has incorporated a kind of circularity into its method. Such circularity is then used as a tool to fashion new and more accurate concepts, which themselves might contain an element of circularity. Kempthorne has this to say on the role of circularity in science:

> The circularity of the scientific method is perhaps not so widely recognized as it should be. This may be brought out by noting that it is only on the basis of hypotheses and theories that the scientist knows what to observe. Facts and theories are intimately intermingled and cannot be logically separated. That is to say, a fact is such only on the basis of some theory, and vice-versa. This statement holds very obviously in the physical sciences; in the social sciences the connection is not always clear but careful analysis shows it to exist.[16]

---

[15]Lashley, K.S. The mechanism of vision: XV. J. gen. Psychol., 1938, 123-193.

[16]Kempthorne, O. The design and analysis of experiments. New York: John Wiley & Sons, 1952.

If Lashley were to be labeled as sympathetic to one school of psychology or another, it would be the gestalt school that would receive his support. The gestalt school is well known for its appeals to common sense, for its many almost inescapable a priori assumptions. Animals are assumed to be thinking creatures, and they display cognition and insight in problem situations. What, then, could be more of a common sense assumption than that animals perceive physically related stimuli as psychologically continuous? And if this stretches credulity too far, does not the existence of a gradient of generalization point to the fact that animals certainly act as if they recognize the psychological continuity of physically related stimuli?

On careful consideration, the argument between L & W and the SG theorists does not seem to be based on any question of whether or not SG is a genuine phenomenon. Rather, the question reduces to one of which hypothesis is to be adopted until a better one comes along. If L & W's analysis of the action of stimuli is to be adopted, then nothing is anything, stimuli become functionally equivalent, and there is no reason why a dog trained to salivate to a tone should not also salivate to a bell, the slamming of a door, a fluctuation of intensity in room lighting, a muscle twitch, the animal's own digestive noises, a flea bite, or the reading of a Shakespearean sonnet. Somewhere, it seems, connections must be postulated between certain stimuli and certain responses. It is unthinkable that all stimuli are equivalent in their effects.

CHAPTER II

THE LASHLEY AND WADE EXPERIMENT

I. THE LASHLEY JUMPING STAND TECHNIQUE

In the previous chapter, it was pointed out that L & W were willing to rest their case on the results of an experiment that eliminated inattention and the role of threshold measurement. In the following paragraphs, a brief description of the experiment reported by L & W as evidence supporting their assertions will be given, and a summary of replies by Hull, Grice, and Grandine and Harlow will be made. However, before going into the details of L & W's study, it would be best to outline the Lashley jumping stand method of training rats, since this is the chief technique L & W used.

The jumping stand is a device originated by Lashley to teach a rat to discriminate between two or more stimuli. Its principle is simple: a hungry rat is placed on a stand some appropriate distance from the stimuli it must choose between; say, in this instance, two cards, one with an 8cm. white circle on a black ground, and the other with a 5cm. white circle on a black ground. One stimulus, say the 8cm. circle, is considered positive, or correct; while the other stimulus, say the 5cm. circle, is considered negative, or incorrect. Previous to the discrimination situation, the rat has learned that if he jumps in the direction of the cards, he will get food. Now, however, cards are placed upright in his path, and his approach to food is blocked by these cards. One card is secured from the rear, so that it

will not fall if the rat jumps against it (this is called the negative stimulus). Every time the rat jumps at the negative stimulus -- in this case a 5cm. circle -- he will bump his nose and fall into a net placed below the cards. If he jumps to the positive stimulus, the 8cm. circle, it will give way and allow the rat to a dish of food. (See Figure 1, below, for photographs of a pigmented rat on a Lashley jumping stand.) The correct and incorrect stimuli are interchanged on some random schedule to prevent the rat's learning simply the position of the correct stimulus, rather than its formal characteristics.

There are a few finer points in the Lashley method which should be introduced and discussed at this point, since any understanding of what will later be talked about hinges on a familiarity with these points.

Suppose a rat is being trained to jump to an 8cm. circle, but to avoid a 5cm. circle. Let us say, for instance, that on some particular trial the 8cm. circle is on the right and the 5cm. circle (locked from behind and unyielding) is on the left. If the rat jumps to the right, the card gives way, and we say no error was made on that trial, and so proceed to place the cards for the next trial, in accordance with some random schedule. But suppose the rat jumps incorrectly, to the left. An error has obviously been made, but we may then follow one of two procedures:

(1) The cards may be allowed to remain in the position they were when the rat jumped, and the rat replaced on the stand. This allows the rat to correct its previous response, and thus it is called the "corrective method." By the corrective method, it is possible that a rat may make an incorrect jump many times before it jumps correctly.

Figure 1A...Front view of the Lashley stand. The animal is tensed to leap to the left; in the text discussion this is a wrong jump, since it is toward the 5cm. circle. If the animal jumps left, he will bump his nose against the locked card and fall into the cloth net shown in the photo.

FIGURE 1. A TYPICAL LASHLEY JUMPING STAND

Figure 1B...A view of the feeding platform to the rear of the stand. In this case, the animal has jumped correctly (to the 8 cm. circle), and the positive card yielded, allowing the animal to the dish of food. The short, black board behind the food dish is propped against the negative card, thus holding it securely in place.

(2) When the rat jumps incorrectly, the cards may be switched according to the random schedule previously set up, and only then is the rat replaced on the stand. For any given trial, then, a rat may jump incorrectly once and only once. Because the rat is not allowed to correct his jump after having missed a previous jump, this procedure is called the "non-corrective method."

From a consideration of the differing natures of the above methods, there arises the question: "What is an error?" Before answering this important question, we must first know what is meant by a trial. The meaning of a trial differs somewhat depending upon whether the corrective or the non-corrective method is being talked about. If the non-corrective method is being considered, then a trial is begun by the placement of the cards, and is ended when the rat jumps, no matter correctly or incorrectly. If the corrective method is being considered, a trial begins also when the cards are placed, but ends only when the rat jumps correctly.

From these definitions of "trial," it can be seen that under the non-corrective method, the animal can make only one misjump per trial, and thus as many misjumps in any given experimental session as there are trials. In the corrective method, however, the number of misjumps that an animal can make on any given trial is limited only by the endurance of the animal and the patience of the experimenter. Consequently, the number of wrong jumps that can be made in an experimental session depends upon whether the corrective or the non-corrective method is used.

The question, "What is an error?", now returns. In the non-corrective method, which allows only one jump per trial, the meaning of an error seems rather clear: a misjump is an error. However, in the corrective method, the meaning of an error is not so clear. It will be

remembered that in the corrective method, it is possible for an animal to jump incorrectly two, three or almost any number of times on any given trial before it finally jumps correctly. If on any one trial a rat jumps incorrectly more than once -- say three times -- are we to say that the animal made one error on that trial, counting all the wrong jumps as one error; or are we to count each incorrect jump as an error, and say that the rat made not one, but three, errors on that trial?

These questions are not merely examples of semantic querulousness, because the answers we accept to the questions could have an important bearing on the conclusions drawn about the existence or non-existence of SG. As a matter of fact, a serious criticism of the L & W study has been made by Grice, and the question of the meaning of an error is central to his criticism.

What does Lashley, the originator of the jumping stand method, consider an error? Lashley says, "By an error is meant ... a trial in which the animal jumped one or more times to the negative card before choosing the positive."[1] This definition of "error" limits the number of errors to the number of trials in an experimental session. Lashley justifies this procedure of counting errors by stating that "We do not know the relative influence of success and error in learning, so any method of counting scores is largely arbitrary."[2]

---

[1]Lashley, K.S. op. cit., p. 130. [2]Ibid.

II. THE L & W EXPERIMENT IN DETAIL

In the following sections, a description of the L & W experiment will be given, and then a summary of experimental and speculative replies by Grice, Grandine and Harlow, and Hull will be presented. The experiment upon which L & W depend for verification of their argument is stated clearly by them. L & W write:

> A simple test of [stimulus generalization] consists of training a group of subjects in a reaction to a single stimulus, then opposing that stimulus to another on the same stimulus dimension and comparing the rates of formation of a discriminative habit when the reaction to the initial stimulus is reinforced and when it is extinguished by the differential training. The experiments [in L & W's article] follow this plan and meet the requirements for a test of irradiation in other respects.[3]

L & W's subjects were rats, whether pigmented or albino is not reported. The stimuli were white circles of 5 and 8 cms. diameter on black grounds. The apparatus was the Lashley jumping stand; a photograph of a typical Lashley jumping stand is shown in Figure 1 on page 14. The procedure of training was the one originated by Lashley.[4] A procedure close to Lashley's own is reported later in this paper, in Chapter III, under "Method, Apparatus and Procedure."

The main L & W experiment (more than one is reported) is ingenious in its approach. Each rat goes through two stages; the first stage is called <u>training trials</u>, or simply <u>training</u>, and the second stage is what shall be called <u>discrimination trials</u>, or simply <u>discrimination</u>. In training, animals are trained to react in some way to a <u>single</u>

---

[3] L & W, <u>op. cit.</u>, p.75.

[4] Lashley, <u>op. cit.</u>

stimulus, a white circle. There are two such ways of reacting: either to jump toward the stimulus, or to jump away from it. Animals trained to jump to the stimulus receive food if they jump to it; they bump their nose on a black card if they jump away from the stimulus. This type of reaction -- jumping toward a white circle stimulus -- is called a facilitation response. If the white circle is made negative (locked), and the black card positive, then the animal must learn to avoid the white circle and jump to the black card. This avoidance type of reaction is called an inhibition response. Since two sizes of stimuli are being used, and there are two ways of reacting to each stimulus, there are four possible conditions for the training series. These conditions shall be designated as follows:

5+ vs. BL- (5cm. circle positive; BLack card negative)
5- vs. BL+ (5cm. circle negative; BLack card positive)
8+ vs. BL- (8cm. circle positive; BLack card negative)
8- vs. BL+ (8cm. circle negative; BLack card positive)

Any given animal was assigned to one such training condition, and one condition only, and was maintained on that condition until he jumped 200 times consecutively without an error.

After the animal was trained to this rigid criterion of learning to a single stimulus, he was advanced to the discrimination trials. In these discrimination trials, he was faced with two stimuli: the original circle in his training and another, different-sized circle of either 5 or 8cms., depending upon which stimulus was used in training. In the discrimination trials, one of two procedures was followed and maintained through discrimination:

(1) the original reaction was reinforced; that is, the animal for whom the 5cm. circle was positive during training found that the 5cm. circle was still positive, while the 8cm. was negative.

(2) the original reaction was extinguished; that is, the animal for whom the 5cm. circle was positive during training found that the 5cm. circle was negative and the 8cm. circle was positive.

We can now designate animals by the sequence of training-discrimination trials they went through. An animal designated as 5+:5- (read "five plus to five minus") is one which was trained to jump to the 5cm. circle during single stimulus training, but had to learn to jump away from the 5cm. circle during discrimination trials. We also call this type of animal an "extinction rat," because its original reaction is being extinguished. An 8+:8+ rat is one whose original reaction to the 8 cm. circle is being reinforced, since the rat still finds the 8cm. circle positive when it begins discrimination trials. Further, for this 8+:8+ rat, we can say that the condition is a "reinforcement of a facilitation," since the animal's original approach response to a circle is being reinforced; while for a 5-:5- rat, we would say that this is a reinforcement of an inhibition, since the original avoidance reaction is being reinforced.

It can be seen that, with two stimuli, two possible modes of response to these stimuli, and two ways of further treating these responses, there are 8 possible conditions in the experiment. Put another way, since each of the four training conditions can be followed by two discrimination conditions, 8 different courses can be followed in the experiment:

REINFORCEMENT	EXTINCTION
(1) 5+:5+	(5) 5+:5-
(2) 5-:5-	(6) 5-:5+
(3) 8+:8+	(7) 8+:8-
(4) 8-:8-	(8) 8-:8+

Unfortunately, L & W did not include all 8 conditions in their experiment. They included only the following: 8+:8+; 5+:5-; 5-:5+; and 8-:8-. Thus, they had in their design the reinforcement of a facilitation, the extinction of a facilitation, the extinction of an inhibition and the reinforcement of an inhibition. While these four conditions control adequately for the relative influence of inhibition and facilitation, they do not account for absolute size preferences rats may have. Both reinforcement conditions operate on an 8cm. circle, while both extinction conditions operate on a 5cm. circle.

To summarize, the situation is as follows. One group of animals was trained intensively to respond to a single stimulus in an unique manner; this group of animals was now placed in a situation where it had to discriminate between the stimulus to which it had reacted during training and a new stimulus. This group of animals found that its original reaction was still correct; thus, the group is called a reinforcement group, because its original reaction is being reinforced.

A second group of animals, trained in exactly the same manner as the first group, found during discrimination that its original reaction was now wrong; it had to change its original response, and learn to react to or away from the new stimulus. This group is called an extinction group, because its original reaction is being extinguished.

Here is the reason for this rather elaborate and ingenious

design. L & W reason, and probably rightly since none of their critics objected to their reasoning, that maximum SG should occur to the stimulus to which the animal was originally trained. If SG is actually operating, the group of rats for which the original response is being extinguished should make significantly more errors than the group for which the original reaction is being reinforced. For example, suppose two rats are trained to jump toward an 8cm. circle, and to avoid the black card. During discrimination, one of these rats will find that it must still jump to the 8cm. circle to get food, while the other rat finds that it must now erase its original response toward the 8cm. circle, and must now avoid it in favor of the 5cm. circle. If the animal whose original response is being reinforced makes N errors during learning, then the animal whose original reaction is being extinguished, and who must unlearn its original reaction as well as learn a new one, should make some number of errors greater than N, say N + Q errors. The extinction animal's tendency to respond to the original stimulus, although such a response is now incorrect, should interfere with and significantly retard its learning of a new and different positive stimulus, IF SG is operating as an influence.

Conversely, if SG is not operating at all in this situation, then there should be no significant difference between reinforcement and extinction animals.

III. LASHLEY AND WADE'S RESULTS

What results did L & W actually get when they conducted their experiment? For four rats, tested on the conditions outlined above,

the results were as follows:[4]

	TRIALS	ERRORS
Original reaction reinforced	175	56
Original reaction extinguished	105	37

In short, just the opposite of what would have been expected if SG were actually operating. The results would have been cause enough for questioning the SG hypothesis, had scores for trials or errors been simply equal.

Similar results were obtained by L & W for four more rats treated in the following manner:

> ...Four rats were trained to choose a white cross opposed to a black card and were given 200 trials of overtraining, averaging 1 error each in the 200 trials. Two of the animals were next trained to choose the original cross and avoid an X of equal size; the other two were trained to choose the X and avoid the original cross.[5]

L & W go on to say:

> The results of these two tests are typical of eight others dealing with surface brightness, direction of lines, number, and various similarities of form, which it is unnecessary to report in detail. In every case a differential reaction was established more quickly when the training involved extinction of the initial reaction to a single stimulus than when that reaction was reinforced. The differences are statistically unreliable but are consistent in all 10 experiments.[6]

L & W report similarly constructed experiments conducted with monkeys and chimpanzees. Results for monkeys fall in line with the rat results; the chimpanzee results favor SG.

---

[4] L & W, op. cit., p.76. [5] Ibid. [6] Ibid.

> Larger numbers of trials and errors [for chimps] were required when the original reaction was extinguished than when it was reinforced but the differences ... are entirely unreliable.[7]

IV. CRITICISMS OF THE L & W EXPERIMENT

The first and still most thorough critique and answer of the L & W study came from Hull.[8] Hull's reply is mostly speculative. His experimental evidence, not including the classical studies he cites, suffers from the very same flaw from which other alleged repetitions of the L & W study suffer: the use of apparatus and procedure far different from any L & W used. However, Hull's speculative arguments are not to be ignored. He writes:

> (A) Lashley and Wade's experiments yield results which are <u>opposite</u> [italics in the original] to what is held to be the implication of stimulus generalization by a falling gradient or even a horizontal gradient.
>
> (B) All ten experiments agree in this reverse outcome....
>
> Now from a statistical point of view it is quite true that the single toss of a coin does not decide anything any more than does a single statistically unreliable experiment. However, in case a coin is tossed ten times in succession with the same result in each case, it is truly remarkable; the probability that this would occur by pure chance is $(1/2)^{10}$, which is but one chance in 1,024. Taking this [L & W's] report at face value, then, it indicates that a generalization gradient of high statistical reliability was really operating in the experiment but that its slope was in the opposite direction from that implied by the stimulus generalization experiments. Thus Lashley and Wade's experiments have a considerably more revolutionary implication than they have claimed. Moreover, these results flatly contradict Lashley and Wade's own conclusion [quoted above] that no graduated spread of the effects of training is obtained. There purports, on the contrary, to be a genuine spread of the effects of the preliminary training, but in the opposite direction.[9]

---

[7] L & W, <u>op. cit.</u>, p.77.

[8] Hull, C.L. The problem of primary stimulus generalization. <u>Psychol. Rev.</u>, 1947, 54, 120-134.

[9] <u>Ibid.</u>

This is the heart of Hull's argument. He goes on to cite a study by Spence "involving simultaneous discrimination in rats which apply directly to the present issue."[10] However, although this experiment's outcome favors SG, it is quite inapplicable to the present issue, since its apparatus is so different from the one L & W used. Spence's experiment seems only to serve to corroborate the almost hackneyed observation that the results reported out of a laboratory are more likely to be a reflection of the biases prevailing in that laboratory than they are of a good answer to a good question.

A further criticism of the L & W study comes from Grice, who has conducted many experiments the results of which proved favorable to the Spence-Hull behaviorist school.[11] Grice purports to have repeated the L & W experiment. Although his results favor SG, as did Spence's, they are not the results of a repetition of the L & W study, since a discrimination apparatus entirely different from Lashley's jumping stand was used by Grice. Thus, L & W are neither confirmed nor refuted by Grice's studies. However, Grice's speculations about the cause of the L & W results have some relevance to the present study.

The first and perhaps least important of Grice's criticisms has to do with a highly technical point, and is purely a matter of speculation, extremely elusive so far as experimental verification is concerned. He writes:

---

[10]Hull, C.L., op. cit., p.131.

[11]Grice, G.R. The acquisition of a visual discrimination habit following response to a single stimulus. J. exp. Psychol., 1948, 38, 633-42.

> One possibility that should at least be considered is that [the L & W] experiment is not a good test of the generalization hypothesis. If the generalization gradient, although a genuine phenomenon, should be relatively flat for the particular dimension selected for study, the difference between the response tendencies to the reinforced stimulus and those to the other stimuli on the dimension would be small.... [and] Differences in the opposite direction, as obtained by Lashley and Wade, would be obtained frequently through sampling error.[12]

However, Grice recognizes that Hull's application of a non-parametric statistical test to the L & W experiments "indicates that this is probably not the explanation of the data considered here." In addition, the L & W study is considered quite an adequate test of the SG hypothesis, by Grandine and Harlow, who say, "The Lashley and Wade technique ... provides an effective method for analysis of the generalization problem...."[13]

It is informative to note, in considering Grandine and Harlow's study using twelve monkeys and the Wisconsin general test apparatus, that they report:

> Learning of the discrimination was more rapid in the generalization tests continuing the condition of reward of the initial training trials than in the generalization tests reversing the condition of reward of the training. Moreover, the differences between error scores on the first test-trials [italics in the original] under the two conditions of reward were statistically significant for most stimulus pairs.[14]

Grandine and Harlow state that they used the *t* formula for testing the significance of a difference in the means of unrelated measures. Their *t* value for size tests was at the 0.1% confidence level.

---

[12] Grice, G.R. op. cit., p.634.

[13] Grandine, L., & Harlow, H.F. Generalization of the characteristics of a single learned stimulus by monkeys. J. comp. physiol. Psychol., 1948, 41, 327-38.

[14] Ibid., pp.327-28.

Grice's second criticism of the L & W study provided the impetus for the first part of the present experiment. Grice writes:

> Probably the most serious complicating factor is the fact that incorrect responses are punished by locking the negative card and causing the animal to fall into a net. It is well known that punished responses are more quickly eliminated than responses which are simply not rewarded. Presumably, this results from a tendency to avoid the punished response. ... Another factor which makes interpretation of the results difficult is the fact that the trial and error figures reported by Lashley and Wade do not include all of the individual jumps made by the animals. ... one 'error' included one correct jump and one or more incorrect jumps. It is possible that if all individual jumps were reported, there would be no difference between the groups, or the difference could even be in the other direction. In any case, if the [extinction group] made more repetitive errors in the early trials [italics in the original] of the discrimination series than the [reinforcement group], this fact would have been obscured by the method of scoring; but it would have been very important in determining the rates of learning of the two groups.[15]

A further criticism, made by both Hull and Grice, involved the small number of subjects in the L & W study. The trend shown by L & W's ten experiments seems sufficient answer to such criticism.

In all the experiments which purport to be repetitions of the L & W study, no one of the investigators used the Lashley jumping stand method. Spence used light and dark alleyways; Grandine & Harlow used the Wisconsin monkey apparatus, and ran only the monkey section of the L & W study; and Grice used an apparatus of his own design, different statistical techniques and different scoring methods.

An adequate investigation of the validity of L & W's claims would involve the use of the Lashley jumping technique, and as close a replica

---

[15]Grice, G.R. op. cit., p.635.

of the L & W design and methods as it is possible to make on the basis of L & W's report. Larger numbers of subjects would be essential to such a repetition, if any confidence were to be placed in the results.

The following chapter will describe an attempt to incorporate Grice's major criticisms into an experiment that at the same time closely approximates the L & W study.

CHAPTER III

THE PRESENT EXPERIMENT

I. A NON-PUNISHMENT LASHLEY JUMPING STAND, AND ITS FAILURE

In an attempt to account for what Grice calls "probably the most serious complicating factor" of the Lashley jumping technique, a Lashley stand was built with provisions to allow for simple non-reward of wrong responses. Figure 2, below, illustrates how this was managed.

In Figure 2, below, it can be seen that the rat has jumped correctly, and so has reached food. If the rat had jumped incorrectly, the

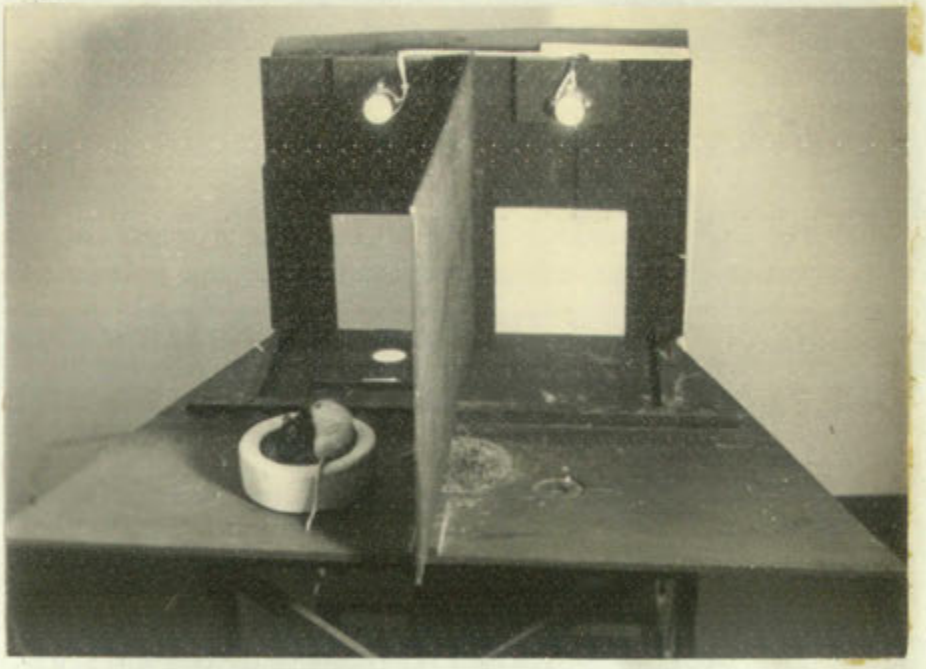


FIGURE 2. A MODIFIED LASHLEY JUMPING STAND
(Food Platform)

negative card would still have yielded, but the partition would have prevented the rat from reaching food. Thus, no punishment was administered the rat for its incorrect response, and learning became a function only of reward or non-reward, without the "complicating" factor of punishment.

Grice's second criticism, concerning the scoring of errors, was accounted for by keeping a complete record of every jump the rat made on each trial. Errors for the day's experimental session of ten trials were broken down into two types: Lashley errors and Grice errors. Lashley errors (abbreviated to "La" on the record form, Figure 3) are simply errors counted by the Lashley system, as described in the last chapter, Page 16. It is possible for an animal to make only as many Lashley errors as there are trials; in this case, there were ten trials per day for each animal. Grice errors (abbreviated on the record form to "Gr") are simply the number of mis-jumps for all trials in an experimental session; the number of Grice errors possible in an experimental session is practically unlimited. The record form on the next page (Figure 3), with a few day's scores for a typical animal, exemplifies the manner in which records of performance were kept, and at the same time shows the difference between Lashley and Grice errors. (The appearance of "L" and "R" in each square indicates the position -- Left or Right -- of the correct stimulus for that particular trial.)

The plan was to run 16 rats on a non-punishment method, but in all other particulars to reproduce the L & W experiment; then it was planned to run 16 more rats on the locking, punishment technique,

LASHLEY JUMPING STAND DATA SHEET

ANIMAL # F-26 (M or F)

SPOTS

CONDITIONS: This sheet records ~~TRAINING~~ DISCRIMINATION

Training Condition $\frac{5}{-}$ vs. BL.

Discrim. Condition $\frac{5}{-}$ vs. $\frac{8}{+}$

DAY	DATE AND WEIGHT	POSITION, TRIAL # AND TOTAL # OF ERRORS PER SESSION										ERROR	
		1	2	3	4	5	6	7	8	9	10	LA	G
27	7/3 - 152	R II	R I	L III	L I	L II	R III	R I	L I	R I	L I	4	6
28	7/4 - 156	L	R I	R II	L II	R II	L II	L I	L I	R I	R II	5	5
29	7/5 - 160	L	L I	R I	L I	R I	R I	R II	L II	L II	R I	3	3
30	7/6 - 155	L	L I	L I	R I	R I	L I	R I	L I	R I	R I	0	0
31	7/7 - 155	R	L I	R I	R I	L I	L I	L I	R II	R I	L	1	1

CRITERION

Figure 3. Record form with separate space for Lashley (La) and Grice (Gr) errors.

which would then constitute an almost exact replica of the L & W study. If SG were favored by the non-punishment method, but not favored by the original Lashley technique, then there would have been evidence substantiating Grice's hypothesis that punishment obscured the effects of SG. If both techniques yielded results favorable to SG, then L & W would have been refuted on their own ground. And, finally, if neither technique yielded data favorable to the Hull-Spence hypothesis, L & W would have been corroborated in a substantial manner. (The fourth possible outcome -- non-punishment favoring Lashley and Wade and punishment favoring SG -- would indicate a rather contradictory state of affairs).

However, the plan never came to completion, since the new apparatus and non-punishment technique did not work. Here is a brief rundown, in a qualitative manner, of how a group of four pilot rats reacted to the non-punishment technique, with further suggestions on how this technique might finally be made to work. Of the four rats trained by the non-punishment technique in the pilot study, two learned the necessary reaction to the single stimulus, while two rats showed no signs of learning. The stimuli used were white circles on black grounds. It was then decided to try a different type of stimulus -- black circles on white grounds, instead of white circles on black grounds. There is some indication that black circles on white grounds will work better on a non-punishment technique than the original stimuli described. There is no apparent reason why this should be, but animals did suddenly begin learning the single stimulus response under the non-punishment conditions when these new stimuli were presented to them. One pilot

rat began to show improved performance when he was deprived not only of food, but of water as well. This may be another possibility to follow up if it is ever desired to make the non-punishment technique workable.

Many of these changes in the technique of training animals were made on the spur of the moment, and not very many days elapsed between the beginning of one treatment and the beginning of another, so that any conclusions drawn about the relative efficacy of varying types of treatments is subject to the criticism that the animals were in the process of "learning how to learn," to use Harlow's term, and that any learning that occurred would have occurred regardless of any changes in deprivation time, stimuli or the nature of deprivation.

II. A REPETITION OF THE L & W EXPERIMENT

After the failure of the non-punishment technique, the next step was to reproduce the original L & W experiment as closely as possible. Twenty-four subjects, pigmented rats, were considered sufficient to yield data that would confirm or refute the L & W results.

Two Lashley jumping stands were used in this study. One was an old stand which had given reliable service for many years, built by a student of Lashley following the pattern of Lashley's first form discrimination apparatus. Another stand was built by the writer in order to make the non-punishment technique a little easier to administer. This new stand was built from an undimensioned drawing in Munn.[1] It

---

[1]Munn, N.L. Handbook of psychological research on the rat. New York: Houghton Mifflin, 1950, p.118.

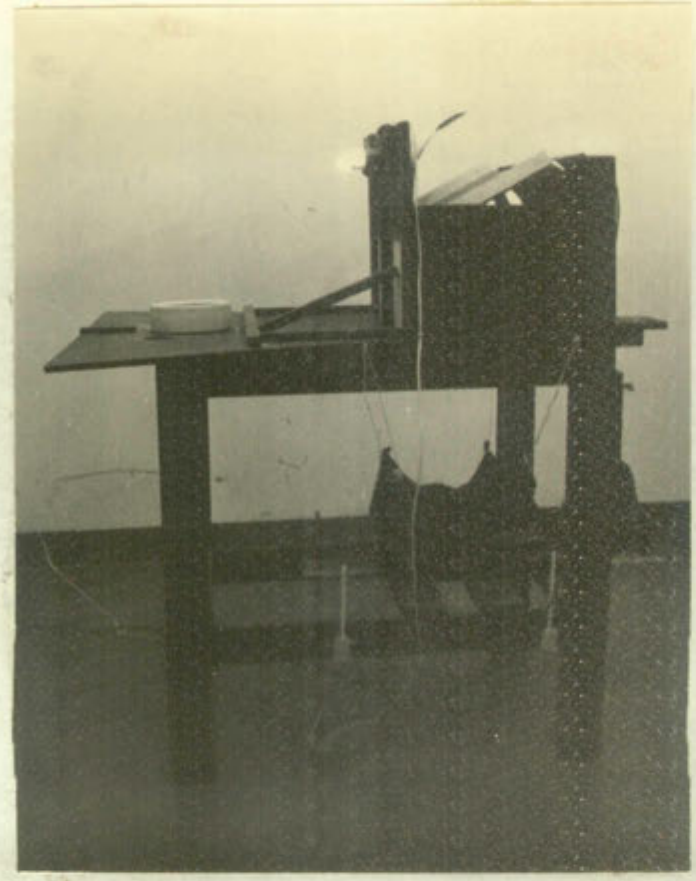
was assumed, without a pilot study, that the punishment technique (the original Lashley technique of locking the incorrect card), would certainly work on this new stand, and so a block of 8 rats was assigned to it for the L & W repetition.

A formal description of the apparatus, method and procedure follows.

III. APPARATUS, METHOD AND PROCEDURE

New Stand -- Figure 4 on the following page is a side view photograph of the new stand and a detail of the jumping platform, net and stimulus cards (note that the stimulus card on the right has fallen, allowing access to the feeding platform). Lights were mounted on this stand to provide uniform and specifiable illumination for the stimulus cards. Lights were also mounted behind the cards to illuminate the feeding area of the stand. The entire apparatus was painted a flat black. The two front lights (the ones lighting the stimuli) were 25-watt frosted bulbs, and were each 11 inches from the plane of the stimuli. The two rear lights were 7-1/2-watt internally coated bulbs, yielding a diffuse, white light.

Stimulus cards were made of heavy white illustration board. The same flat black paint with which the entire stand was painted was used on these stimulus cards to enclose the area of white which acted as the stimulus. The centers of stimuli were 10 inches apart, plus or minus about a quarter of an inch, depending upon the placement of the cards. Cards were practically flush with the discrimination area, and there was no place for the rat to grasp in case he jumped at the incorrect



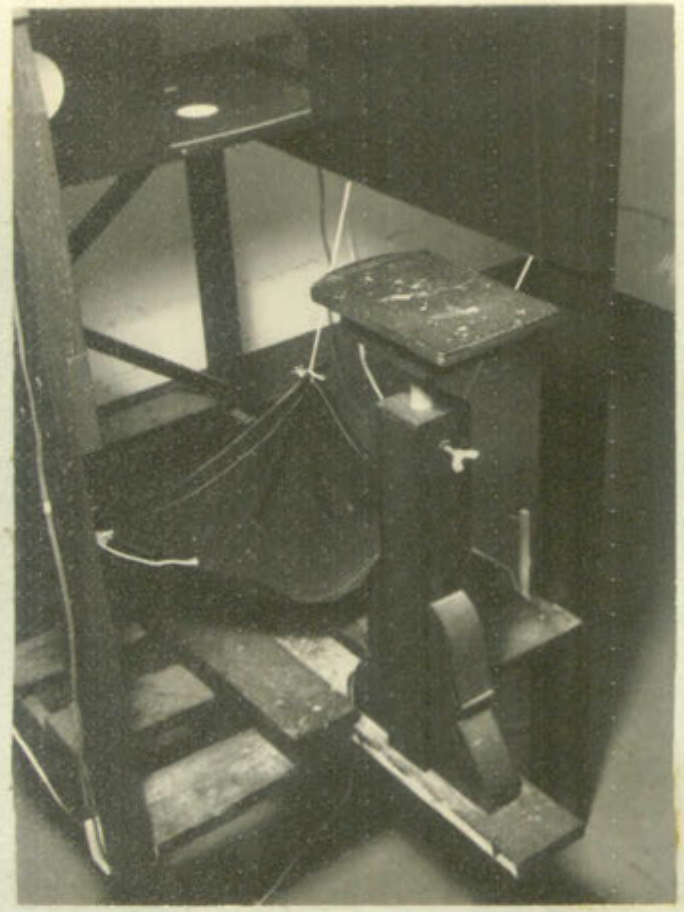


Figure 4. The new stand -- a modified Lashley jumping stand. The photo to the left is a side view of the stand, showing the feeding platform. The photo on the right is an enlarged detail of the jumping platform, net and stimulus cards. Partition is not up for the punishment method.

stimulus. Such grasping, on other stands, keeps the rat from falling. The fall to the net from the center of the stimuli was approximately 25 inches. The area of the jumping platform was approximately 43 square inches, being 5-3/16" by 8-5/16", with a radius of curvature at the front end of approximately 4-1/2". The jumping platform was relatively rigid and free of sway.

Old Stand -- What is here called the "old stand" is one which had been in the animal psychology laboratory for many years; this stand was built by a student of Lashley. Figure 1 on page 14 shows two views of this stand, and it can be best compared with the new stand by comparing Figures 1 and 4. The major differences between the old stand and the new stand are as follows:

	OLD STAND	NEW STAND
Length of fall from stimuli to net	33"	25"
Dimension of jumping platform	6" x 7"	8-5/16" x 5-3/16"
Distance between centers of stimuli	11-3/8"	10"
Stability of jumping platform	Swayed much	Quite rigid
Lighting	Room	Built-in

In most of the essential structural characteristics, the two stands seemed quite similar, at least insofar as purpose is concerned.

Both stands were used in the same room. This room was a windowless room, so that lighting was comparatively even for all groups of rats run on any particular stand. Rats were run in the evening in the case of hoods, during the day in the case of albinos. Temperatures in the room hovered around 80° F. (estimated) and 25% humidity (also

estimated). The experimental room contained a fan, which provided a subdued but constant source of noise.

The following is an exact reproduction of the protocol that was consulted and followed for all groups of rats. Individual differences in rats required some departures from this protocol, in order to accustom them to their task. The major difference in treatment between rats was in the matter of coaxing and urging. Some rats required such extra treatment, while others did not. Here is the protocol as it was followed throughout the experiment:

LASHLEY JUMPING STAND PROCEDURE --- Lashley Method

1st Day: 4 rats on stand, all at once, 1/2-hour exploration. Jumping platform moved out gradually until animals must make small jump. Animals replaced frequently. No partition used, large food dish in center of feeding platform.

2nd Day: 10 free-choice jumps to food. Two rats run at a time, alternately. Gradual moving out of platform to 16 inches, which seems to be a good general distance.

Since some of these animals will not jump unless urged, at least in the first 8 or 10 days of the training period, a rough time limit of 30 seconds is placed on them before they are forced to jump. If an animal shows signs of jumping by himself, he is allowed to jump, since it is considered better just to let him jump in the hope that he will get the habit of jumping without urging. Such voluntary jumping is to be desired over continual urging, because urging can sometimes be disruptive.

3rd Day: The stimulus card which the animal must avoid is placed in position and locked. Ten jumps away from this card by Gellerman's orders, one rat at a time.[2]

---

[2] Gellerman, L.W. Chance orders of alternating stimuli in visual discrimination experiments. J. genet. Psychol., 1933, 42, 207-208.

4th Day: Same as 3rd Day.

5th Day: Both cards now in place, and training to single stimulus is begun.

Rats are kept on roughly a 24-hour deprivation schedule. "Roughly" because each group of four rats are rotated, the rat which was last of the group to be run on one night being the first of the group to be run on the next night, and the second of the group on the night after that. This keeps deprivation time, popularly (in the psychological literature) equated with "motivation," reasonably even.

During training to the single stimulus, the maximum number of wrong jumps allowed per trial, before guiding, was five. If an animal jumps incorrectly to a given position on a given trial five times in a row, he is then guided and forced to the correct side. Generally, longer feeding is allowed after an error has been made. .

The training series consists of 200 jumps (20 days) not necessarily errorless.

When rats have completed 200 training jumps, they are then confronted with two circles, only one of which is correct.

Rats are considered to have learned the discrimination when they jump correctly 19 out of 20 jumps.

To start with, subjects were 24 hooded rats (pigmented) and, later on, 8 albinos. Since it was not foreseen that the new stand would not work, the 8 hooded rats that were being run on it had to be discarded and not included in the experimental results. These 8 rats would not learn at all adequately even the simplest task -- jumping to a single stimulus -- and so it was considered carefully and decided that these rats would not constitute either a fair test of the SG hypothesis or even a rough approximation to the L & W study. This left a remainder of 16 rats running on the old, more efficient stand. Of these 16 rats, 3 or 4 either would not jump or would not sufficiently learn the response to the single stimulus. Since the thorough training of the rat to the single stimulus is central to the rationale of L & W's study,

these rats had to be discarded, also. What was left comprised 8 rats, four extinction and four reinforcement, but due to an error that occurred during the filling of an experimental block, one condition was repeated. Thus, the factorial design outlined on a previous page was not complete, although there was still a group of four reinforcement rats to compare in performance against a group of four extinction rats. Because 8 rats did not seem sufficient to lead to any even partially valid conclusions, another group of 8 rats, these being albinos, was selected from out of a group of about 22. These rats were selected not on the basis of how they discriminated, but merely on the basis of willingness to jump and ability to learn the simple single stimulus task.

All rats reported on (16 in all) were run on the old stand, except where stands are being compared; in this case, the data for 8 hoods on the new stand are compared against the data for 8 hoods on the old stand.

Since there was no known way to equate groups on ability to learn a single stimulus task, statistical controls were run. Within strains, animals were found to be about equal in ability to learn; in any case, no significant differences were found, within strains, between the ability of the reinforcement group to learn the training task and the ability of the extinction group to learn the same task.

CHAPTER IV

RESULTS, DISCUSSION AND CONCLUSIONS

I. A NOTE ON THE STATISTICAL TREATMENT OF THE DATA

L & W do not report the type of statistical analysis they employed in determining whether or not their data yield significant differences. Grice, on the other hand, in discussing the treatment of his data, states:

> The skewness of the distributions and the non-homogeniety [sic] of the variances make the t-test inappropriate for testing the significance of ... differences. Consequently, non-parametric tests of significance were applied.[1]

Grandine and Harlow make no statements about whether or not their data fulfill the requirements that indicate the use of techniques that assume normality of data, and they state that they employed the t-test for differences between unrelated means.[2]

However, the samples in this study were so small that no adequate test of the normality of the data could be found to determine if parametric tests would apply to them. A persistent difficulty in the application of non-parametric tests to any learning data is the rather rigid requirement of non-parametric statistics that the outcome of each event be completely independent of the outcome of another event. To use the analogy of the coin, even if you were to throw 100 heads in a row, the probability of throwing a head on the 101st throw remains 1/2, or one out of two. The coin did not learn, by past experience, the way living

---

[1] Grice, op. cit., p.638. [2] Grandine and Harlow, loc. cit.

organisms do.

But if a rat jumps in some particular direction on his very first jump, the direction of his subsequent jumps may possibly depend upon his past experience with the jumping situation. This possibility cannot be denied, and the non-parametric method used by Grice is inapplicable to the data obtained in the present study.

If both the F-test and non-parametric techniques could not be used without violating one or another rule concerning their use, what course could be taken? What was decided in this case was to accept a possibility, rather than to flout a certainty. Since there was still some possibility that the data were normal, even though no known test could have shown this for the small samples involved, it was considered a safer course to use the F-test on the data and not to use the available non-parametric techniques. Thus all conclusions drawn and levels of confidence reported are subject to the blanket modification, "... if the F-test is applicable to this data."

Actually, the situation is not quite as hopeless as it first appears. It is true that nothing is known about the form of the distribution from which the scores come on the small samples involved in this study; but Fisher has this to say about the use of the F-test on non-normal distributions: "... [it has been] demonstrated how closely the theoretical distribution was verified in material that was far from normal."[3]

Early in the statistical analysis of the data, it was found that

---

[3]Fisher, R.A. The design of experiments. New York: 1951, Hafner, 6th Ed.

any comparisons made for both Grice and Lashley errors always yielded lower Fs for Grice errors than for Lashley errors. This was attributed to the relatively larger variability of Grice errors. Thus, in the following report on results, when "errors" is the measure being used, only Fs derived from Lashley errors will be reported.

II. RESULTS AND DISCUSSION

Table I is a summary of the performance of the 16 rats (8 hoods and 8 albinos) which were run on the old stand. Since the 8 hoods run on the new stand never even learned the single stimulus task, their data is not included, except where they are compared against the 8 hoods on the old stand to demonstrate a difference between stands. Since the answer to the major question with which this experiment deals hinges on the performance differences, or lack of differences, between the reinforcement group and the extinction group, the data are arranged so that these performances can be readily examined and compared.

Table II provides the data from which were plotted the curves in Figures 5 through 8. This table offers a convenient comparison between the reinforcement and the extinction group for any given strain by juxtaposing the mean number of errors made by the reinforcement group on a particular day and the mean number of errors made by the extinction group on that same day. This is done for discrimination days 1 through 10. Since there were two types of errors that could have been made, error scores are recorded under their appropriate heading -- Lashley errors or Grice errors. Figures 5 through 8 are graphic representations

TABLE I. TRIALS AND ERRORS
DURING DISCRIMINATION FOR 8 HOODS AND 8 ALBINOS[1]

HOODS

REINFORCEMENT GROUP

RAT #	CONDITION	TOTAL TRIALS	ERRORS LA	ERRORS GR
F-1	5-:5-	49	13	13
F-2	8+:8+	93	13	16
F-3	5-:5-	75	11	14
F-4	8-:8-	219	90	123
	TOTAL:	436	127	166

EXTINCTION GROUP

RAT #	CONDITION	TOTAL TRIALS	ERRORS LA	ERRORS GR
F-5	5+:5-	178	55	68
F-6	8+:8-	254	115	167
F-7*	5-:5+	400	145	209
F-8	8-:8+	230	53	70
	TOTAL:	1,062	368	514

ALBINOS

REINFORCEMENT GROUP

RAT #	CONDITION	TOTAL TRIALS	ERRORS LA	ERRORS GR
M-9*	5+:5+	250	154	340
M-10	8+:8+	119	36	78
M-11*	5-:5-	260	74	88
M-12	8-:8-	55	13	17
	TOTAL:	684	277	523

EXTINCTION GROUP

RAT #	CONDITION	TOTAL TRIALS	ERRORS LA	ERRORS GR
M-13	5+:5-	59	11	16
M-14*	8+:8-	240	141	387
M-15	5-:5+	238	84	125
M-16	8-:8+	113	42	82
	TOTAL:	650	278	610

*Had not reached criterion when experiment was terminated.

[1]All values are figured up to, and including, the criterion.

of these data. There are two graphs (two separate pages) for each strain; one graph depicts mean Lashley errors for the reinforcement and extinction groups, while the other graph shows Grice errors for the two groups.

Table III provides data for a comparison of the two stands. Training error scores are given for 8 hoods on the new stand as well as for 8 hoods on the old stand.

Since different strains of rats were used in this study, it was considered necessary to search for possible differences between strains. Table IV provides data for such a comparison, although it does so in an indirect way. Table IV lists Lashley and Grice error scores during training for those animals which had to learn an approach response and for those animals which had to learn an avoidance response. In a previous chapter, these responses were labeled facilitation and inhibition responses. These data are shown for both hoods and albinos.

Statistical analysis of the data.

When total errors was used as the measure for testing differences (see "ERRORS" column in Table I), the F-ratio for reinforcement hoods vs. extinction hoods was 4.05; this figure approaches the 5.99 required for the 5% confidence level, and since one extinction animal was still running and making errors when the experiment had to be terminated, it is possible that this F-ratio might have come even closer to the 5% level of confidence. This result is in a direction unfavorable to L & W. Since two reinforcement albinos and one extinction albino never finished, total errors could not be used as a measure in comparing albinos.

TABLE II. MEAN ERRORS PER DAY[2] FOR FIRST TEN DAYS FOR FOUR REINFORCEMENT RATS VS. FOUR EXTINCTION RATS

	HOODS					ALBINOS				
	REINF.		EXT.			REINF.		EXT.	
DAY	La	Gr	La	Gr	DAY	La	Gr	La	Gr
1	4.5	10.2	5.7	19.2	1	2.3	3.3	4.6	14.0
2	3.5	3.7	6.0	12.0	2	3.3	3.3	5.6	13.0
3	3.2	4.5	4.7	8.7	3	4.6	8.3	4.6	7.3
4	2.0	2.2	4.0	5.0	4	4.3	10.3	5.0	5.6
5	2.5	3.5	5.2	8.7	5	3.0	4.0	3.0	3.3
6	3.2	3.7	4.5	6.2	6	2.6	5.3	3.6	5.3
7	1.7	2.0	4.7	5.2	7	3.0	4.3	1.6	2.0
8	1.7	1.7	5.0	6.2	8	3.0	5.3	4.3	6.0
9	1.5	1.7	3.2	4.2	9	3.0	3.3	2.6	3.3
10	2.0	2.0	3.7	3.7	10	2.0	2.0	1.6	2.3

TABLE III. RAW DATA FOR COMPARISON OF STANDS. ERROR SCORES DURING TRAINING FOR 8 HOODS ON OLD STAND VS. 8 HOODS ON NEW STAND.

	8 HOODS ON OLD STAND				8 HOODS ON NEW STAND			
RAT	TRAINING	ERRORS		RAT	TRAINING	ERRORS	
#	CONDITION	La	Gr	#	CONDITION	La	Gr
1	5-	8	15	17	5-	35	91
2	8+	24	43	18	8+	29	53
3	5-	19	28	19	5-	39	87
4	8-	26	46	20	8-	27	38
5	5+	27	70	21	5+	37	78
6	8+	17	38	22	8+	22	38
7	5-	13	15	23	5+	63	80
8	8-	10	14	24	8-	34	64
		144	269			286	529

[2] A rat that finished in less than 10 days was presumed to have learned to the point where, if it had continued to be run, it would have made no more than 1 Lashley and 1 Grice error per night; in computing means, such rats were carried on this basis.

TABLE IV. FACILITATION VS. INHIBITION DURING TRAINING FOR HOODS AND ALBINOS

HOODS

FACILITATION				INHIBITION*			
RAT #	TRAINING CONDITION	ERRORS La	ERRORS Gr	RAT #	TRAINING CONDITION	ERRORS La	ERRORS Gr
F-5	5+	27	70	F-1	5-	8	15
F-6	8+	17	38	F-7	5-	13	15
F-2	8+	24	43	F-3	5-	19	28
	TOTAL:	68	151	F-4	8-	26	46
				F-8	8-	10	14
						76	118

ALBINOS

FACILITATION				INHIBITION			
RAT #	TRAINING CONDITION	ERRORS La	ERRORS Gr	RAT #	TRAINING CONDITION	ERRORS La	ERRORS Gr
M-13	5+	14	40	M-15	5-	38	78
M-14	8+	3	3	M-16	8-	33	81
M-9	5+	10	16	M-11	5-	40	80
M-10	8+	3	5	M-12	8-	46	80
	TOTAL:	30	64		TOTAL:	157	319

*Due to an error, this block of animals contained 5 inhibition and 3 facilitation rats; the proper proportion should have been 4 and 4. F-ratios were computed using formula for unequal Ns.

When total trials to criterion (see Table I) was used as a comparative measure, hoods yielded an F between the reinforcement and extinction groups of 6.64, well within the 5% confidence level (5.99 was required). This result favors Grice, Hull, et al, and is unfavorable to the L & W interpretation.

The F-ratios obtained on hoods tend to favor the SG hypothesis, although no striking across-the-board significance was apparent.

Figures 5 through 8 graphically display the relationships within the data in Table II; in addition, these graphs provide some empirical basis for a couple of the intuitive statements that are made toward the end of this paper.

In Figures 5 and 6, graphs of the hood data, the extinction group is at every point higher than the reinforcement group. It is intuitively very difficult to look at such wide differences and still to maintain that there is no difference between the groups; it is even more difficult to state that these differences might have occurred purely by chance since, if non-parametric statistics were applicable here, this could occur only once in 1,024 times.

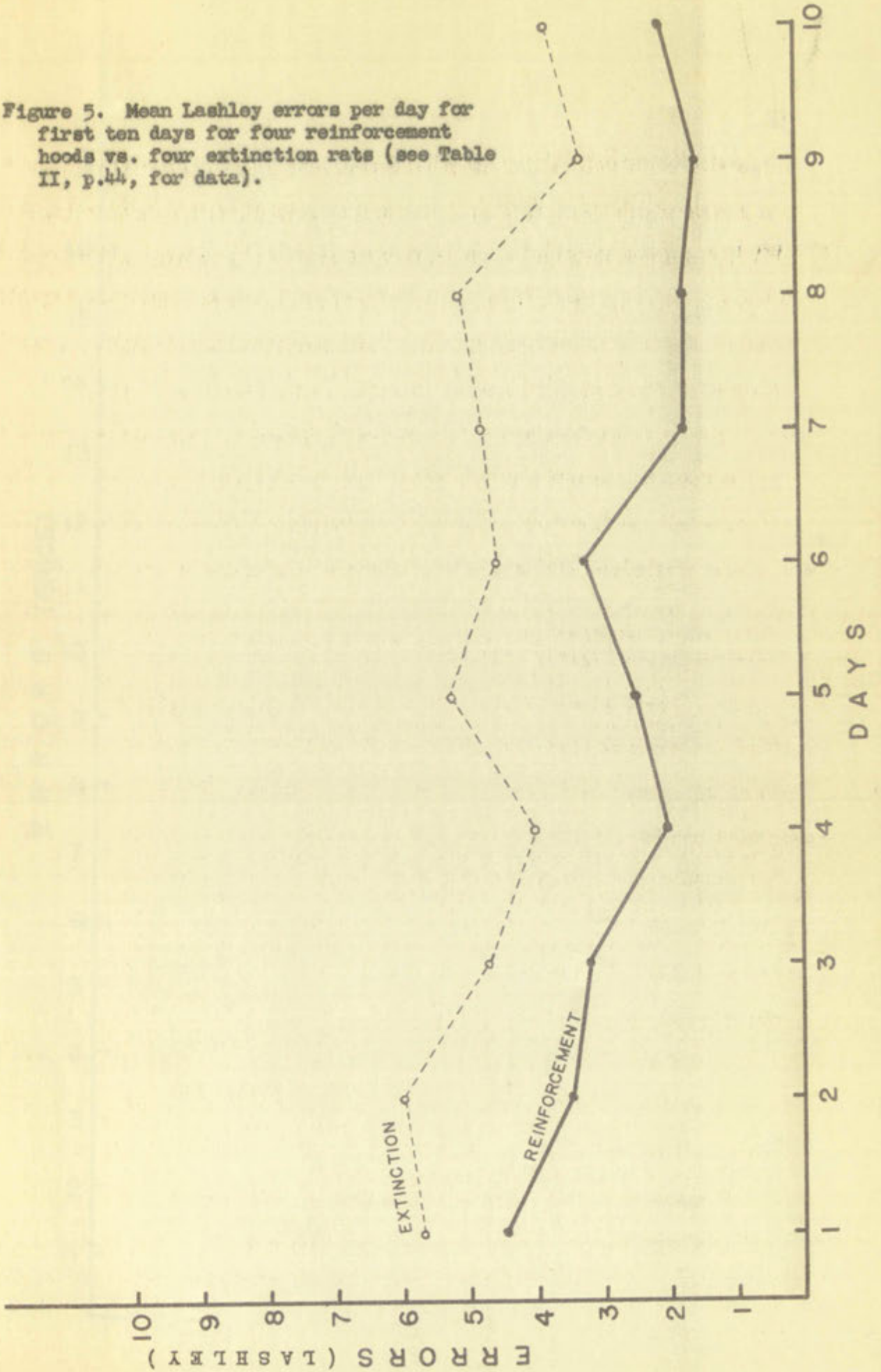
The story for Figures 7 and 8 is quite a bit different. These graphs depict Lashley and Grice errors between reinforcement and extinction groups for albinos. Here, no consistent differences are found, and the lines for reinforcement and extinction animals criss-cross apparently at random. If, in the light of the results that were obtained with albinos, these graphs do not argue very strongly for the superiority of reinforcement over extinction animals, they do make a

strong argument for a possible difference in strains (when viewed in contrast to the results for hoods) which, in any previous or future repetition of the L & W study, should be considered as a determining factor in the nature of the experimental results.

An interesting by-product of the experiment arises in the comparison of strains. No significant difference between strains was found when errors or trials were used as comparative measures; however, since there were two experimenters, one running hoods and the other running albinos, there is a confounding of the influence of experimenters and of strains, so that no very strong conclusion should be drawn on the basis of error or trial measures about the lack of difference between strains. It was found further and in a most strikingly significant degree, that there was some strain difference with respect to a preference for jumping toward or away from a stimulus in a single stimulus learning task. It was found that, with albinos, total errors during training for facilitation animals -- those who had to jump toward the single stimulus to get food -- were significantly less than total errors during training for inhibition animals -- those who had to jump away from the single stimulus to get food. The computed F was 68.74, while the F required for the 1% confidence level was 13.74. For hoods, there was no significant difference between inhibition and facilitation animals. A strain difference seems to be indicated from the above results; this strain difference is probably due to the hoods' superior vision.

A significant difference was found between stands, when training

Figure 5. Mean Lashley errors per day for first ten days for four reinforcement hoods vs. four extinction rats (see Table II, p.44, for data).



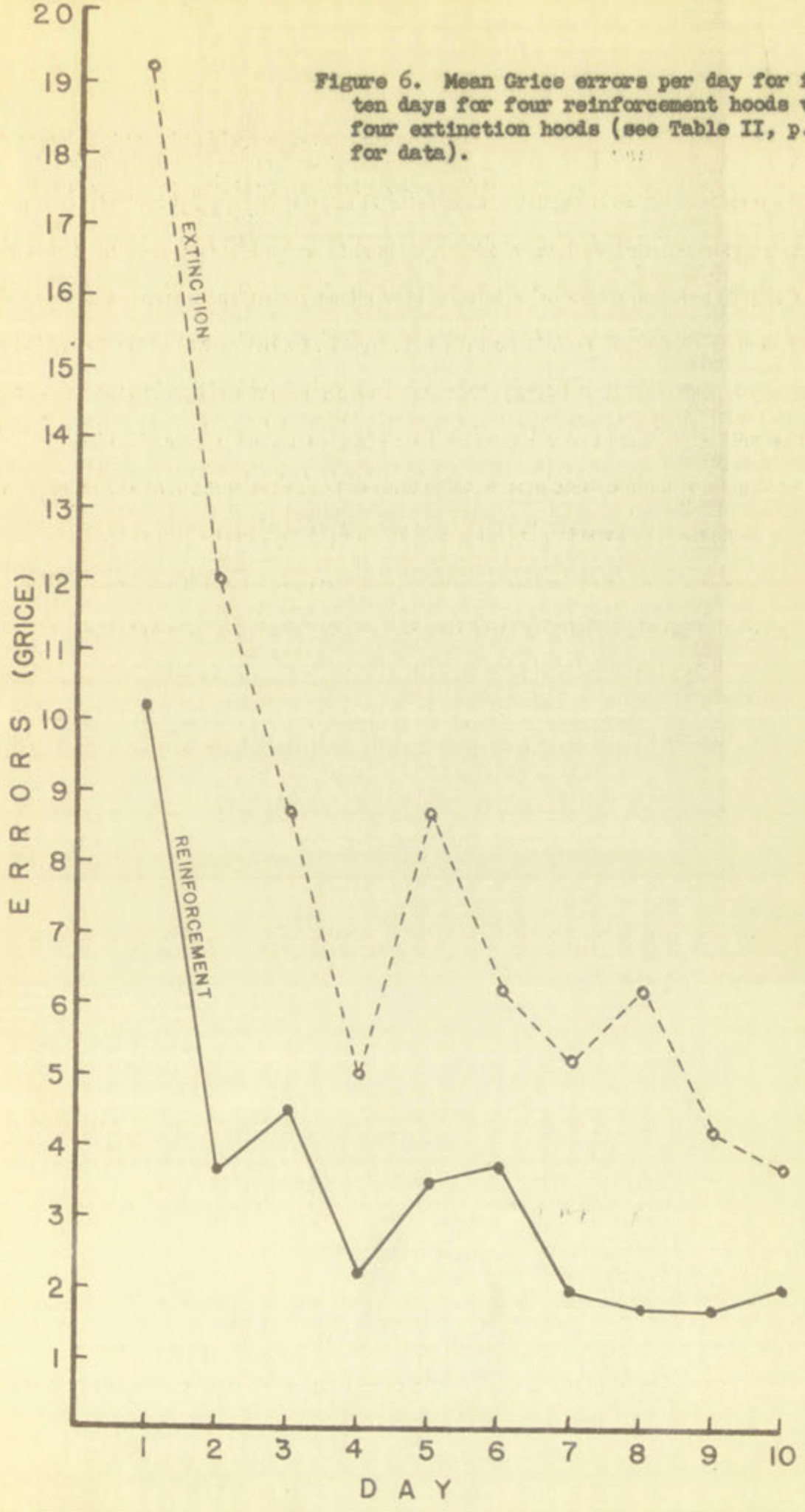


Figure 6. Mean Grice errors per day for first ten days for four reinforcement hoods vs. four extinction hoods (see Table II, p.44, for data).

Figure 7. Mean Lashley errors per day for first ten days for four reinforcement albinos vs. four extinction albinos (see Table II, p.44, for data).

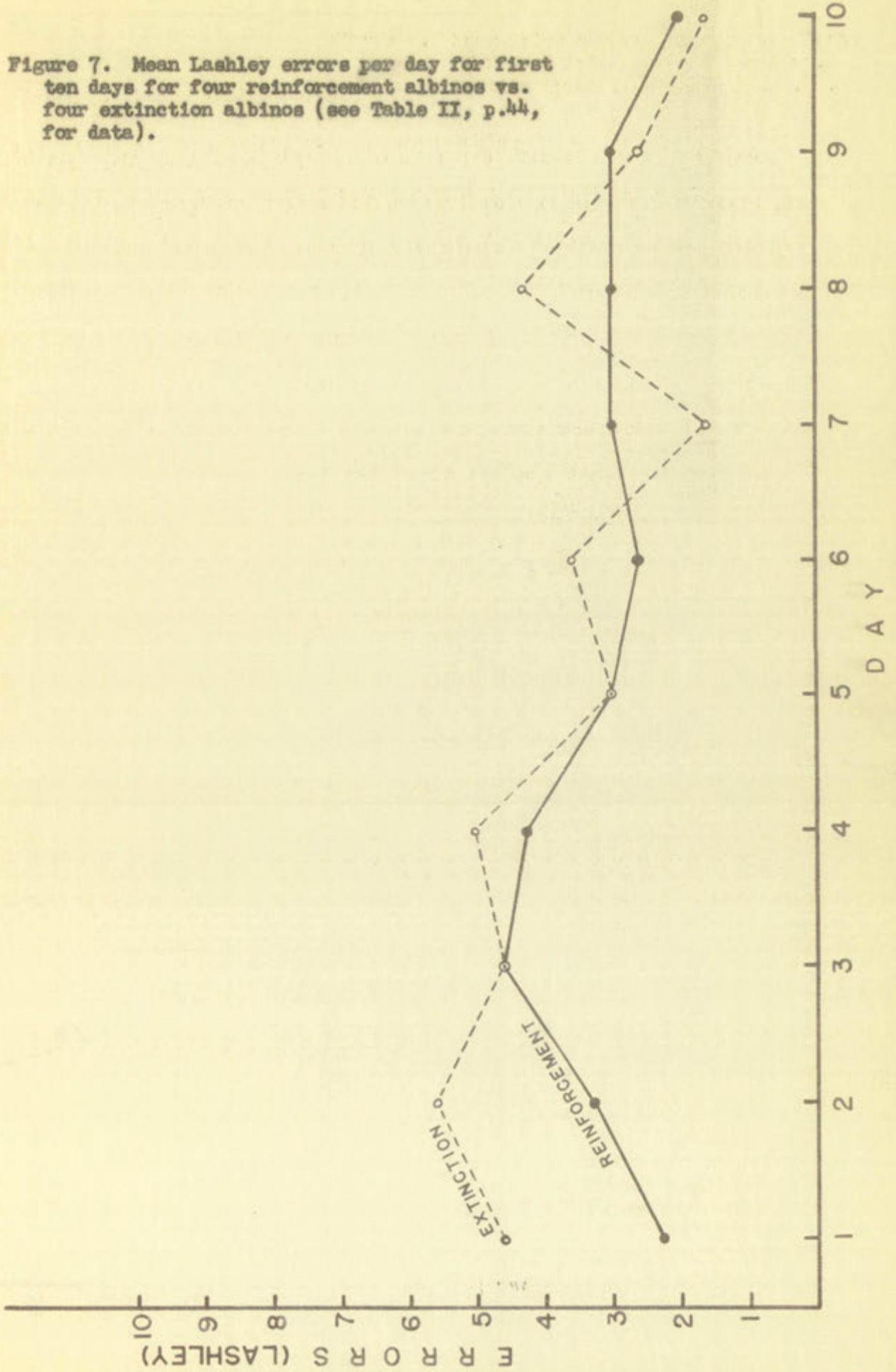
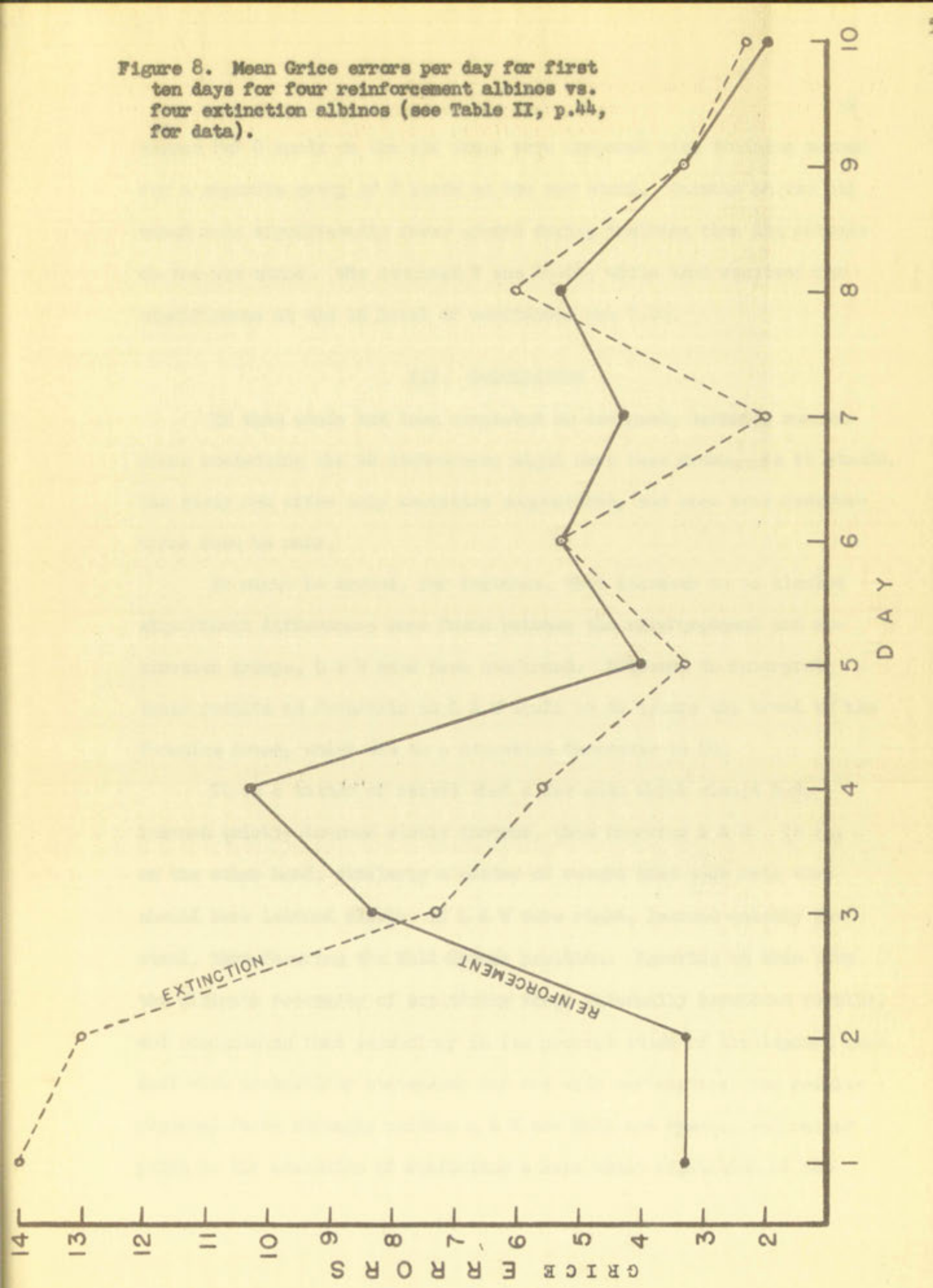


Figure 8. Mean Grice errors per day for first ten days for four reinforcement albinos vs. four extinction albinos (see Table II, p.44, for data).



scores for 8 hoods on the old stand were compared with training scores for a separate group of 8 hoods on the new stand. Animals on the old stand made significantly fewer errors during training than did animals on the new stand. The computed F was 14.04, while that required for significance at the 1% level of confidence was 8.86.

III. CONCLUSIONS

If this study had been completed as designed, definite conclusions concerning the SG controversy might have been drawn. As it stands, the study can offer only tentative suggestions, and even here reservations must be made.

It might be argued, for instance, that inasmuch as no blanket significant differences were found between the reinforcement and extinction groups, L & W have been confirmed. However, to interpret these results as favorable to L & W would be to ignore the trend of the F-ratios found, which was in a direction favorable to SG.

It is a matter of record that a few rats which should have learned quickly learned slowly instead, thus favoring L & W. It is, on the other hand, similarly a matter of record that some rats that should have learned slowly, if L & W were right, learned quickly instead, thus favoring the Hull-Spence position. Ignoring at this time the ultimate necessity of explaining these internally anomalous results, and recognizing that psychology in its present stage of development must deal with probability statements and not with certainties, the results reported favor strongly neither L & W nor Hull and Spence, but rather point to the necessity of conducting a more exact repetition of the

L & W study with a much larger number of subjects. It has been further suggested that a more crucial test of generalization would entail training large numbers of animals to the necessary criterion, and then recording only the very first jump of the discrimination trials; this would allow the use of a non-parametric test of significance and would eliminate the objection of dependency.

It is the feeling of the writer that if this experiment were repeated as briefly outlined above, the results would favor the SG position, and not the position taken by Lashley and Wade.

The differences in apparatus and subjects found should serve as a warning to any animal experimenters undertaking to investigate the results of other workers in the field, that apparently unimportant differences in procedure can lead to the drawing of conclusions different from what might have been expected.


DIETER
BINDS
BETTER