AN EMPIRICAL STUDY OF THE EFFECTS OF ANXIETY AND
FAILURE-SUCCESS UPON THE PERFORMANCE OF A COMPLEX TASK

DISSERTATION

Presented in Partial Fulfillment of the Requirements
for the Degree Doctor of Philosophy in the
Graduate School of the Ohio State
University

By

JEAN LUCILLE BURTON, B. A., M. A.

The Ohio State University
1956

Approved by:

[Signature]

Adviser
Department of Psychology
# Table of Contents

Introduction ........................................... 1

Review of Literature and
Statement of the Problem ............................ 6

Methodology ............................................. 24

Results .................................................... 36

Conclusions .............................................. 42

An Addendum ............................................. 47

Bibliography ............................................. 51
List of Tables

I. Table of Mean Accuracy Scores .................. 37

II. Table of Significant t-tests ..................... 38

III. Table of U-test comparisons of
differences between trials ....................... 39

IV. Table of U-test comparisons of
differences between groups ...................... 40

V. Table of Re-organization of the original
experimental groups ............................ 49
INTRODUCTION

This paper will present an attempt to explore the question of the effect of anxiety and failure-success upon performance. The decision to use a personality variable, i.e., anxiety, and a learning-experimental variable, i.e., failure-success, within the same experiment stemmed from the following considerations:

First, the present day clinical psychologist who has even minimal acquaintance with general personality and learning theory soon finds himself in the uncomfortable position of attempting to reconcile the often marked differences between his observations of individual behavior and his expectancies for behavior based on the results of theoretical research. These difficulties seem to arise from several obvious sources. On the one hand, many of the clinical concepts deal with deviant behavior and are based on theoretical assumptions which are explanatory rather than predictive. On the other hand, the more adequately formulated theories, such as the Hullian view, which do allow for prediction and experimentation, place the emphasis upon development of general principles of learning.

Second, most personality theories utilized in clinical practice (psychoanalysis and its derivatives) have arisen from observation of
the extremes of behavior -- behavior that might be considered to be $\pm 2$ or $3\sigma$'s from the mean in any distribution. On the other hand, most of the experimentally-based approaches, particularly those in the learning area, deal with the "average", or behavior falling $\pm 1\sigma$ around the mean in any distribution. Thus, the clinician has available a fair amount of predictive information about "average" behavior. However, when such information is utilized in the attempt to predict the behavior of deviant individuals, problems may arise. In addition to these above-mentioned difficulties, there is a major problem inherent in attempts to generalize and apply principles derived from different population samples.

Third, generally speaking, personality theorists have been concerned with the study of those aspects of behavior which remain relatively consistent over a wide range of situations. Learning theory has, on the other hand, been primarily concerned with situational factors which produce consistencies in behavior over a wide range of subjects. It would seem that a study of the interaction of these two "kinds" of variables might offer the possibility of a rapprochement between personality theory and learning principles.

The three points just discussed are problems that have long been recognized. However, only recently has there been any concentrated effort to deal with them. There are several areas illustrative of this effort, such as the recent experimental work on rigidity, on the "authoritarian personality", and on anxiety. It is the latter concept which the dissertation utilizes, as recent experimental research centering on "anxiety", has reformulated this concept so that it may be
dealt with experimentally and still retain its clinical usefulness. "Anxiety" has numerous and diverse meanings both to the layman and to the social scientist. In the language of the "man-on-the-street" this word may be used as an expression of concern as to whether the local bus will be on time, or, as a description of the reactions of a person preparing to undergo surgery. Its use by social scientists shows parallel extremes. It may range from the status of a basic theoretical concept, as in psychoanalysis, to a label encompassing various discrete overt behaviors, as used commonly in clinical practice. In the past, the publications dealing with "anxiety" have been as diversified as the apparent meaning of the word itself. These have covered the continuum from "pure" experimental studies of the animal laboratory to textbooks devoted entirely to discussion of the meaning of the term in clinical application (13, 20, 29, 32, 37, 41). Despite the conflicting interpretations, anxiety, as a concept, has remained deeply entrenched in psychological vocabulary and in psychological thinking.

Within the past few years, however, a somewhat new approach both to the concept itself and to its study has been formulated. It is an approach which attempts to place a generally accepted clinical or behavioral description of anxiety into an experimentally testable theoretical framework. This approach may be exemplified by the work of Taylor and her associates (48, 49, 50, 51, 52, 53, 54), who have dealt with anxiety as a motivational variable and thus as a theoretical concept which is measurable through a range of reported behaviors. Thus, the concept can be perceived as on a continuum of "amount of" or "de-
gree of the behavior defining it. The operational quantification of this concept by Taylor has enabled experimenters to investigate it in the same manner in which they have approached other concepts of motivation, i.e., with the assumption that anxiety is "present" or "exists" in all individuals to some degree, at some time.

Anxiety cannot be controlled by the experimenter in quite the same way that such motivators as hunger or thirst can be. It is a variable which is "inherent" in the individual in the same way as intelligence or mechanical aptitude is "inherent". It may be provoked or enhanced by an experimental condition, but may not necessarily be controlled by the experimenter as to degree. Neither is it necessarily oriented towards a specific, known goal. Further, as it is operationally defined as differing quantitatively within individuals, its effect in terms of its drive aspects can be approached best through the comparison of differing groups within an average population, i.e., through the study of individual differences.

The utilization of Taylor's approach to anxiety has resulted in considerable research, which, despite the diversity of orientation of the experimenters, is consistent and communicable. From this research, general problems have become apparent which seem to warrant further investigation. Some of these are: how does anxiety affect learning? What are the differences in effects of anxiety upon tasks of varying complexity? What is the effect, if any, (on performance) of the interaction of anxiety and other personality variables? What is the effect, if any, of the interaction of anxiety with external variables, such as the nature and complexity of the task?
These are the kinds of questions that have contributed to the formulation of this dissertation. Specifically, the problem to be considered here is what, if any, are the effects of the interaction of a personality variable, anxiety, with a learning variable, failure-success upon performance.

The choice of failure-success as the learning or experimental variable was based on the following considerations: first, it is an experience that all individuals continually encounter and the individual difficulties of dealing with this experience make it an important clinical problem; second, there is often a marked difference between what is observed with the individual case and the effects of failure and success expected on the basis of theoretical or normative data; and, third, since, in our culture, failure-success is an accepted form of social motivation, it might logically be expected to have a significant relationship to a personality variable such as anxiety.
REVIEW OF THE LITERATURE AND STATEMENT OF THE PROBLEM

Review of the literature pertaining to the Taylor Manifest Anxiety Scale.

The development, reliability and validity of the scale

There is a generally recognized set of overt behaviors which clinicians label "manifest anxiety". The agreement about the behaviors which make up this pattern is independent of the differing theoretical implications inferred from the behaviors. Taylor (53) has taken this common agreement as the basis for the development of The Manifest Anxiety Scale.

Taylor had five clinicians judge the Minnesota Multiphasic Personality Inventory items on the basis of whether or not the items reflected "manifest anxiety". Cameron's (7) description of chronic anxiety reactions was the criterion for "manifest anxiety" (53). The 65 items on which there was at least 80% inter-judge agreement were then placed among buffer items. The latter had also been taken from the MMPI but had been judged as non-indicative of anxiety. The scale was then administered to 352 beginning psychology students. Fifty of the items which had the highest correlation with the total anxiety score were then retained and placed among 175 buffer items (53).

Taylor (53) then tested the scale for: (a) Reliability -- after
a three week period a reliability coefficient of .89 was obtained (N of 59). On a test-retest over a five to seventeen month period, the correlations obtained were .82 and .81 (N of 163). (b) Validity -- the scale was administered to 1971 beginning psychology students; the mean anxiety score was 11.56 (out of a possible 50), with 20% of the group obtaining scores above 21. The comparison group consisted of 103 diagnosed neurotic and psychotic individuals; the median score for this group was approximately 34 (50). Taylor concluded that the results "appear to indicate that there is some relationship between the anxiety-scale and clinical observation of manifest anxiety" (53, p.290). The range obtained and the mean differences between the groups would seem to support the assumption of at least some validity for the instrument.

Other studies that investigated the instrument itself are relatively few. On the positive side, Westrop (56) found a relationship between four signs in the Rorschach Anxiety Indices and the TMAS scores. Holtzman, Calvin and Bitterman (19) reported a correlation of .86 between the TMAS and Wenn's "Scale of Neuroticism", and on test-retest found no practice effects or sex differences.

Sampson and Bindra (144) compared neurotics (clinically rated as anxious and non-anxious) and normals. Although the range of scores was the same, the mean of the neurotic group was 26, compared to 16 for the normal group. No relationship was found between the clinical rating of the neurotic as "anxious" or "non-anxious" and the scale scores. However, at a statistically significant level, the anxious patients scored between 19 and 33, while the "non-anxious" neurotics
scored either above and below this. The authors suggest that the Taylor scale does not measure degrees of manifest anxiety and should be used only as a separation measure of individuals who are likely to be classified as anxious by clinical criteria.

Negative results were reported by Bitterman and Holtzmann (5) who found no relationship between anxiety as measured by the Taylor scale, and anxiety as measured by the Rorschach and as inferred from clinical interview. On a complex perceptual task, Meyer, Bahrick and Fitts (33) reported that the Taylor scale was a less sensitive indicator of disturbance than the Rotter Incomplete Sentence Blank and, thus, did not differentiate subjects in relation to their performance as well as did the Incomplete Sentence Blank scores. Although the Hilgard, Jones and Kaplan (18) research was not specifically designed as a validity study, their results are pertinent to this topic. They repeated, with some variation, Taylor's original eyelid conditioning experiment (this study will be discussed in the next section) and failed to find a significant correlation between the anxiety score and the rate of conditioning. It should be pointed out that the last three studies cited used the total range of Taylor scores, whereas the majority of the other studies used some method of selecting subjects from the upper and the lower range of scores.

The conclusions that can be drawn, at this point, about the TMAS are that: (a) while it may not include all behaviors which are labeled "anxious", it does bear enough relationship to clinical diagnosis that the scale will differentiate the anxious from the non-anxious; (b) the instrument is not particularly successful in establishing a con-
tinuums of degree of anxiety and, when so used, the results are equivocal; (c) at this stage, the only "proper" use of the scale is one which utilizes it to separate subjects on some basis of extremes of scores. Only when so used does the instrument have reliability and validity.

The scale as a method of investigating the effects of anxiety on learning.

Taylor (52) and Spence & Taylor (51) did the two "classical" experiments in this area. They investigated the relationship between anxiety and simple eyelid conditioning. The conditioning of the anxious subjects was significantly superior to the non-anxious. These results were interpreted in terms of Hullian learning theory, i.e., "manifest anxiety" reflects variations in generalized drive levels" (51, p.183).

The Hullian formulation postulates that any increase in D (the total effective drive strength operating at a given time, which includes both relevant and irrelevant drives) would increase the probability of evocation of all responses operant in a given situation. Thus, when there is a single or an extremely dominant response tendency, an increase in D should lead to better performance. When many competing response tendencies are present, the effect of increased drive would be dependent upon the relationship between the original strengths of the correct and incorrect response tendencies.

With regard to the last point, Farber and Spence (11) found that the anxious subjects did significantly better on a conditioning task (single response tendency) and significantly poorer on a maze task.
(many competing response tendencies) than did the non-anxious. Their major conclusion was that any increase in drive will affect performance, but the specific effect will be dependent upon the nature of the task. The same problem was investigated by Montague (36). Nonsense syllables of different associative value were used. The assumption made was that the larger the number of associations to any syllable, the greater would be the number of potential incorrect response tendencies. The results were in accord with theoretical predictions. The anxious subjects performed less well when having nonsense syllables of high associative value, and did better than the non-anxious when learning nonsense syllables of low associative value.

Ramond (39) attempted through pre-training to control the number and strength of competing responses. He paired stimuli to arbitrary responses, choosing the latter so that varying numbers of response tendencies would be aroused. When many competing response tendencies were present in the learning situation, the anxious subjects, as predicted, did more poorly than the non-anxious. However, there was no difference between the groups when the learning task had minimal competitive response tendencies, although the anxious would be predicted to perform better than the non-anxious. Two possible explanations were offered. The first was based on stimulus generalization, i.e., that the incorrect response generalized to the stimulus word and thus gained strength. The second explanation was that the higher drive level of the anxious subject made for potentially greater excitatory strength of incorrect responses, which in turn, affected the oscillation threshold, and, thus, allowed for the occasional dominance of the
incorrect responses.

The last study to be reported, that of Lucas (28), is one of the very few which has attempted to explore the interactive effects of anxiety and another variable within a learning theory framework. The design was the learning of intra-serial duplication lists by anxious and non-anxious subjects who received differing amounts of failure. Both groups were expected to decrease in performance with increased failure and with increased intra-serial duplication. On similar lists, the anxious subjects were predicted to make a significantly greater decrease than the non-anxious. However, the non-anxious subjects showed a constant improvement in performance with increase in failure. Lucas, in discussing the results, states that there is a need "for some new theoretical assumptions regarding the properties of anxiety and failure" (28, p.65). He suggests the following:

An obvious assumption is that the responses produced by anxiety and failure increase at a rate different from those produced by the corresponding drives. Suppose, as the level of tension is increased, the increase in the production of interfering responses is at first slower and then more rapid than the increase in drive. There would then be an optimal level of tension at which the increment in drive is greatest relative to the increment in the response. In the present experiment, the interpretation would be that the non-anxious Ss were approaching such a level, and the anxious Ss had passed it and were on the ensuing decline" (28, p.65).

From the studies reported in this section, one can conclude, tentatively, that the effect of anxiety can be predicted within the framework of the Hullian concept of drive (a) when relatively simple learning tasks are involved and (b) when anxiety is the only variable under consideration. Thus, it can be seen that the effect of anxiety is dependent upon the complexity of the task and upon its interactive effect
with other variables.

The scale as a method of investigating the relationship between anxiety and other experimental variables

The limited number of studies found in the literature which attempted to relate anxiety and perceptual theory reported similar results. A case in point is the work of Bitter and Kniffin (6) in which some of the previous perceptual studies utilizing neutral and taboo words were repeated. They found no significant difference between anxious and non-anxious subjects although there was a significantly higher reaction time for both groups attached to taboo words.

One of the few studies reported on the relationship between anxiety and another personality variable was Wesley's research (55). She compared "rigid" (defined by scores on a scale developed from the MMPI), "anxious" (defined by scores on the Taylor scale), and "normal" subjects on a concept formation test similar to the Wisconsin Card Sorting Test. Although there were no differences between the groups in number of trials to reach criterion, the rigid group took longer to shift set and gave more perseveration responses. One of the conclusions was that there appeared to be no relationship between the personality variables of rigidity and anxiety.

An investigation of the effects of anxiety in a life situation was attempted by Sarason and Mandler (45) using a questionnaire they developed, similar to the TMAS. The subjects were first dichotomized on the basis of the upper and the lower 30% of the scores obtained on the scale; each of these groups was further sub-divided into upper and lower 50%. The subjects predicted their own grade averages (PGA) for
that quarter; the actual grade average (AGA) for each subject was obtained at the end of the quarter. The predictions for the two high anxious groups were that the higher the anxiety score, the lower would be the PGA stated by the subjects, but, the higher would be the AGA achieved by the subject. The exact reverse predictions were made for the two low anxious groups. Significant differences were found in the predicted direction between the combined high anxious groups compared to the combined low anxious groups, and, between the two sub-divisions of the low anxious. However, no differences were found between the two sub-divisions of the high anxious. The conclusion was that "the low end of the anxiety distribution approximates a behavioral continuum, while the high end does not" (45, p.815). The results and conclusion presented some problem in interpretation which the authors attempted to solve by stating that anxiety is not only a function of the strength of drive, but, of (a) the adequacy of the defense of the individual, and, (b) the correlation between the drive and the choice of the defense.

Although the authors may have been justified in stating that drive alone is not an adequate concept with which to predict the effects of anxiety, their explanation seemed somewhat strained. The explanation was post hoc, not particularly related to their experimental design and was phrased in rather untestable terms. It would appear from previous studies cited that the anxiety scales are not extremely precise or sensitive instruments. Thus, any attempt to sub-divide groups, other than on the basis of extremes, would probably duplicate the same confusing results obtained in the above study.
There have been a few studies which investigated the relationship between anxiety and the various forms of external motivation usually subsumed under the label "stress". Because of their pertinence to this dissertation they shall be reported in some detail.

Westrop (56) studied the performance on a digit symbol task given under "stress" conditions, which included shock, the effect of being observed, etc. The stress conditions apparently impaired performance generally but not differentially in relation to high or low anxiety. Sarason, Mandler and Craighill (146) also used the digit symbol test and gave the subjects a verbal report of "expected" or "not expected to finish" at the beginning of the task. There was no difference between the high and low anxious group without stress ("not expected to finish"), but the low anxious did significantly better than the high anxious under stress ("expected to finish"). Although not statistically significant, the high anxious subjects tended to perform at a higher level under the non-stress than under stress, and the low anxious subjects tended to do better under stress than non-stress.

Bitterman and Holzmann (5) studied conditioning and extinction of the GSR in a stress situation. They reported no relationship between changes in behavior under stress and anxiety as defined by the TMS scores.

Mandler and Sarson (31) used success, failure, and neutral reports about performance on Kohs Blocks and Digit Symbol. They reported that tasks at their best level when they were subject to neither success nor failure, whereas the low anxious subjects did best when given a failure report.
A final illustrative study was that of Deese, Lazarus, and Keenan (10). They studied the performance of high and low anxious subjects on a standard nonsense syllable task under a variety of stress conditions. There were three groups: a control group which received neither threat nor punishment; an avoidance-learning group in which the subjects were given shock either when they failed to respond or when they gave an incorrect response; and, a non-avoidance-learning group who were shocked as often for correct as for incorrect responses. Under the avoidance conditions, the anxious did significantly better than the non-anxious. The same direction of differences, although not statistically significant, was found with the control group and with the non-avoidance conditions. In analyzing the sources of the differences, they concluded that, "the large differences in the avoidance group are due both to a decrement (in performance) of the non-anxious subjects, and a facilitation (of performance) in the anxious subjects; in the non-avoidance group, the difference is entirely due to a decrement (in performance) in the non-anxious....the non-anxious subjects had a decrement in performance under stress condition while the anxious had some improvement when the stress was such that it could be avoided by them....these (results) suggest that an anxiety drive explanation is too simple." (10, p.58)

The conclusions from this section are: (a) anxiety as a motivational variable has somewhat selective drive functions, i.e., although present (or "inherent") in the subjects, the presence or absence of its effect upon performance may be a function of other variables; personality, task, etc.; (b) although there are differing results, there is
a consistent interactive effect between anxiety and "stress"; (c) the interaction seems to be, in part, a function of the type of stress used, i.e., if one viewed this along a "reality continuum", the more experimentally contrived stress (as shock) produces less detrimental effect in anxious subjects than does the use of a more customary stress (such as failure); (d) the interactive effect of the two variables, at this point, cannot be explained completely.

Review of the literature pertaining to success and failure experiments.

Verbal "failure" and "success" have had a long and somewhat undistinguished history of usefulness in psychology. These are two "workhorse" variables often used by psychologists as a means to an end, but rarely perceived as worthy or study in and of themselves. Historically, verbal failure was simply used as one method of inducing "punishment"; and, success was used to induce "reward". In research on learning, the classical experiments utilizing verbal failure-success stemmed from Thorndike's conceptualization of the Law of Effect, and the controversy arising from it. Stemming from this position of being means of weakening and strengthening "connections", verbal failure-success had a ready-made role in the experiments of modern learning theory—that of a method of investigating negative and positive reinforcement.\(^1\)

\(^1\)There is an area of research which views failure and success from a completely different view—that of information theory. Within such a framework, failure and success are dealt with as sources of information which the subject can utilize to improve subsequent performance. This approach is being acknowledged but not reviewed as this dissertation is concerned with the traditional concept of failure-success.
Within recent years, verbal failure has been subsumed under the general concept of stress while success has been perceived as either a counteractant to stress or, more neutrally, as non-stress. Thus, both nominally have been given a new role in both personality and learning theory — that of inducing forms of motivation. The theoretical importance of this difference in role is essentially in how failure-success is expected to affect performance. Viewed as reinforcement, failure-success would be assumed to act upon the response or habit; whereas, viewed as a motivation-inducer, failure-success experience would be assumed to result in affective change in the subject which in turn would affect subsequent performance. In either role, the general effect of failure could be stated as that which detracts from, interferes with, or decreases performance; of success, that which heightens, aids, or increases performance. These generalized results are predictable from various theoretical views and there is such a tremendous body of research to support them that they have become almost "givens" in psychological knowledge.

However, even casual observation would find many instances in which these principles do not operate as predicted. There is an increasing group of studies investigating these discrepancies. While these specific studies view verbal failure-success as an inducer of motivation, they recognize that the affective reactions of the individual (which may be reflected in his performance) are a function of
the failure-success experience interacting with certain aspects of that individual's personality structure. As a consequence, the effect of failure-success upon performance may be expected to differ from individual to individual qualitatively, as well as quantitatively. If it is assumed that there is a simple generalized affective reaction common to all individuals, the effect of failure-success can be studied in a random sample. If, however, the interaction of the failure-success experience with individual personality aspects is assumed to be important, it is then necessary to choose groups of subjects on the basis of differences between individuals on the given personality variable to be studied. Unfortunately, there does not yet exist a body of theory which can predict with precise accuracy these interaction effects. Before such a body can be developed, considerable empirical research must be undertaken in order to determine which personality aspects actually interact with failure-success, and in what manner. Inasmuch as this research is concerned with these particular problems, the review of the literature shall be restricted to an illustrative sample of studies which are pertinent to these problems, and which indicate the various areas in which empirical research has been done. The studies have been selected on the basis that: (a) there had been some sorting of subjects in terms of some clearly defined personality or intellectual variable, and, (b) while the failure-success administered dealt with the subject's total performance, it was not such as to offer corrective information which could be utilized by the subject on subsequent performance.
Research studying failure-success and personality variables

D. R. Miller (3k) tested the general hypothesis that if symptoms of the various personality disorders are interpreted as defenses against failure, then the differing defenses should result in differential reaction to failure. There was a control group of normal subjects and four experimental groups consisting of conversion hysterics, neurasthenics, paranoid schizophrenics, and those with character disorders. A level of aspiration task was used; the failure-success patterns were predetermined and were the same for all subjects. The author stated that all five groups responded differently to failure and described the results somewhat qualitatively as follows: those classified under character disorders "seemed so fearful of failure that they cannot even risk predicting improvement even when such a prediction is warranted," as in contrast to the normal group who "... when things seemed to be going well, have the courage to risk prediction of substantial increases. In fact, the normals are unique among the groups in their ability to reverse the direction of their prediction when such a reversal seems justified by the situation" (3k, p.386).

Kelman (21) utilized the autokinetic situation to investigate the effects of failure-success on suggestibility (defined as following suggestions of a "confederate" about the autokinetic phenomenon). The subjects were placed in one of four groups -- Success, Failure, Ambiguous Conclusions, and Control. The results from these groups were related to the subjects' scores on Ascendancy, Lack of Inferiority Feelings, and Lack of Nervous Tension (three factors taken from Thurs-
tone's GAMIN). Each group was sub-divided on the basis of the median score on the factors. These comparisons indicated that the high-success group was more suggestible than the low-success group; the reverse relation existed for the remaining three groups. The over-all results found the failure group significantly more suggestible and the success group significantly less suggestible than the control group.

Research studying failure-success and intellectual-achievement variables

One of the best known studies in this area is that by P. S. Sears (17) dealing with children's levels of aspiration. There were three groups of subjects: A failure group which was composed of children who had received past academic failure in reading and arithmetic. A success group which was made-up of those children who were above-average in academic accomplishment. A differential group composed of children who had had academic success in reading but failure in arithmetic. The level of aspiration task measures were reading and arithmetic problems. One-half of each group was given success on the arithmetic and failure on reading, while the other halves received converse experiences. Although the over-all results were that the failure group had significantly higher discrepancy scores and greater variability than the success group, all of the groups showed lower discrepancy scores and less variability when reacting to success than when reacting to failure.

Rosenzweig (42) utilized a wide variety of subjects, (normal, mentally-retarded, and "problem" children, as well as normal, maladjusted
and schizophrenic adults). The subjects were given failure and success experiences on various problem solving tasks and then allowed to work on their own choice of tasks. The normal children and adults preferred to return to the tasks they had failed, the maladjusted and schizophrenic chose the success tasks, and the mentally retarded and problem children randomly split their choices.

Using a somewhat different approach from the other reported studies, Postman and Brown (38) differentiated their groups on the basis of experimentally administered failure and success experiences. They then tested the subjects' perceptual sensitivity to success and failure symbols. On the words which connoted failure, the failure group was significantly more sensitive than the success group; whereas, with the words connoting striving or success, the success group was more sensitive.

Statement of the problem

From the review of the literature it is apparent that there is no one theory or framework that can completely predict or even explain the variety of divergent results of the "anxiety" and/or failure-success experiments reported. It would seem that the framework that would appear to offer the best basis for prediction and/or explanation is the Hullian view, or a modification thereof. Thus, it is from a Hullian framework that the specific hypotheses of this dissertation were drawn, even though the results of other research appeared to suggest that exact and accurate predictions are not always possible.

The formulation of the hypotheses of this research involved the
following conceptualization for the variables of anxiety and failure-success: First, there is an optimal range of general drive state, within which individuals may perform adequately. If, however, the general drive state is increased or decreased beyond this range, the performance of the individual will accordingly decrease. Second, anxiety is construed as an "internal" drive, contributing to an increase in general drive state. Third, failure and success are construed as "external" drives. Failure is perceived of as contributing to an increase in general drive state, while success contributes to a decrease in the general drive state.

Thus, it would follow that the highly anxious individual should have a higher drive level, and the non-anxious individual should have a lower drive level than individuals operating within the optimal range of general drive state. While failure should increase the general drive state for all individuals, its interaction effect with anxiety would be such that the low anxious individual would move closer to the optimal drive level (and therefore, toward improved performance), whereas, the high anxious individual would move further from the optimal drive level, in the direction of decreased performance. The effects of success upon the general drive state, and thus, on performance, would be the exact opposite of the effects of failure.

Within this framework, the hypotheses for an investigation of the interaction effects of anxiety and failure-success upon performance must be as follows:
(a) an anxious group will decrease their performance after a failure experience, and will increase performance after a success experience.

(b) a non-anxious group will increase their performance after a failure experience, and will decrease performance after a success experience.

(c) an anxious group will increase their performance after a success experience, and will decrease performance after a failure experience.

(d) a non-anxious group will decrease their performance after a success experience, and will increase performance after a failure experience.
METHODOLOGY.

The choice of the experimental task was governed by the following criteria: (a) the task should be one of a relatively complex nature, applicable to a college population, but not so complex as to discriminate between higher levels of intelligence; (b) the nature of the task had to be such that verbal failure-success could be given in a manner which was consistently meaningful for the subjects, that is, the task should be of such a nature that later failures or successes would appear to be as effective as earlier ones; (c) the task should be such that marked learning from trial to trial either would not occur in great measure, or could be controlled.

A concept formation test, The Wisconsin Card Sorting Test (WCST) appeared to fit the various above-mentioned criteria (4). This test was developed by Grant and Berg (16) as a modification of the Weigl-Goldstein-Vigotsky concept formation test, and it consists of:

...four stimulus cards and 64 response cards, which are devised so that each card contained from one to four identical figures of a single color. Four kinds of figures were used: stars, crosses, triangles and circles. Four different colors were used: red, yellow, green, and blue. A single card might have four red stars, or two green circles or any of the 64 possible combinations of colors, forms, and numbers. Each card could then be sorted or categorized according to the color, the form,
or the number of the figure. A sorting tray consisting of four double compartments was used. The four stimulus cards put before S in the upper set of compartments from left to right were: one red triangle, two green stars, three yellow crosses, and four blue circles. (16,p.404-405)

The usual administrative procedure for the WCST was to instruct the subject to sort the response cards into four groups, placing them in the appropriate compartments underneath the four stimulus cards. As the subject sorted each card he was told whether he was "right" or "wrong". One of the three dimensions, color, number or form, would be the "right" response. As soon as the subject sorted ten consecutive cards correctly the experimenter "shifted" the correct basis of sorting to another dimension (with no explanation to the subject). This process was repeated until the subject had sorted twice on each of the three dimensions.

Initially, the choice of the WCST as the experimental task appeared warranted by (a) the fact that it seemed to meet the previously specified task-criteria for this dissertation and (b) its utility in other studies.2 (see footnote 2 on the following page) However, during the pre-testing some major difficulties were encountered in its use.

One of the difficulties encountered with the WCST was that the trials needed to reach the criterion for "shift" varied considerably from subject to subject, particularly with the initial dimension of sorting. This gave rise to several possible limitations. First, inter-subject variation in the original learning opportunity could mask experimental effects on the subsequent trials. Second, this variation in initial learning resulted in variable lengths of time needed by the
subjects to complete the experimental task. As only one hour of experimental time was available, the data from any subject not completing the task within this period was lost. Since time was an important limiting variable, unknown and biasing effects could presumably influence the results if this variable were ignored.

Another major difficulty was that the WCST, once utilized, did not appear to meet the third criterion for an acceptable task. It seemed to allow considerable positive transfer, which tended to negate or suppress the measurable effects of the experimental variable on performance. The recognition that this particular problem was, perhaps,

2Grant, Jones and Tallantis (17) used the WCST to test Heidbruder's concept formation hypothesis that the perception of concrete objects was the dominant mode of perception. Contrary to this hypothesis, the results suggested that number concepts were more readily formed than either color or form concepts. Grant (15) repeated the experiment since further investigation indicated that the subjects responded to the configuration formed by the objects rather than to the actual number of objects. With this factor altered, it was found that number concepts were not utilized as readily as color or form concepts. Other studies utilized the instrument to investigate personality variables. Ross, Rupel and Grant (13) studied the effect of distraction, heckling and shock upon "abstract behavior" (defined as performance on the WCST). Shock (alone or in combination with heckling or distraction) was the only variable affecting performance. However, the subjects rapidly regained their previous level of performance after shock. Fey (12) compared the performance of schizophrenics and "normals" on the WCST and the Wechsler Adult Intelligence Scale. There were no significant correlations between success on the WCST and on Wechsler total or subtest scores. However, the schizophrenic group did significantly poorer on the WCST than did the "normal" group. Wesley (55) found that, although not statistically significant, the anxious group did obtain consistently lower scores both in number of trials taken to shift set and in the number of perseverative responses as compared to the normal group.
more a function of the experimental design than of the WCST itself, was quite apparent, but, appeared to be as insoluble as the other difficulties mentioned.

Therefore, a modification of the WCST was developed and then tested on a small number of student-volunteers from introductory psychology courses. The results of this pre-testing indicated that while the difficulties encountered with the WCST had not been completely eliminated, they were sufficiently reduced so that the modified test could be utilized.

The experimental task — a modified card-sorting task

The modification of the WCST had four rather than three dimensions which could be used for the card-sorting. The original three dimensions of the WCST, color, number, and form, were retained and a fourth dimension, "position", was added. Position was defined as the placement of the forms in quadrant location (for example, the upper right corner of the card), and, further, placement of the forms randomly within a given quadrant. The random placement was necessary to avoid having the subjects respond to configurations or patterns of forms as occurred in the experiment conducted by Grant and Jones (see footnote 2).

Each of the four dimensions had, in turn, four variations. Thus, there were one, two, three, or four red, blue, green or yellow circles, squares, triangles or crosses placed in any one of the quadrants.

The number of dimensions and variations allowed for 256 different combinations of form, color, number and position. However, only 225 of these combinations were used in the experiment, as the pre-testing
established this number as that which the average subject could be expected to sort within the time limit. Twelve of the remaining combinations were used for the stimulus or model cards. There were no exact duplications in either the twelve stimulus-combination cards or the 225 response-combination cards (a given card could potentially match another on one, two, or three dimensions but never on all four dimensions).

The principles of the sorting of the cards also differed from the WCST in several important ways: first, the cards were to be sorted on the basis of a combination of two of the four dimensions, rather than on a single dimension; second, the cards were divided into three sets of 75 each and used for three separate tasks rather than for one continuous task; third, the "shift" in sorting occurred after a given number of trials rather than after a given number of consecutive correct sorts; and, fourth, only one shift occurred in each set of trials.

There were six permutations possible when the four dimensions were used in pairs or combinations. This allowed two different sets of combinations of dimensions to be used for each of the three sets of cards. Specifically, Set A (#1-75) was to be correctly sorted by each subject on the combination of color-and-form, which shifted to a position-and-number combination. Set B (#76-150) had first the number-and-color combination and then shifted to position-and-form. Set C (#151-225) was to be correctly sorted first on form-and-number and then on position-and-color.

For each subject, the shift from one sorting combination to the other occurred on the 31st trial of each set of cards. This provided
30 trials on the first and 45 trials on the second combination in each set.

It should be noted that position was always one of the sorting dimensions in the second combination of each set. It had been apparent in the pre-testing that this was the dimension that was most difficult for the subjects to utilize. Therefore, it was placed in the last combination of each set and a disproportionate number of trials was given to the last combination. These two steps were taken in an attempt to increase the range of errors and also to increase the possibilities of successful solution of the task.

There was a possibility, of course, that a particular combination or set of cards could differ from others in difficulty of sorting. In an effort to control this, the order of presentation was varied. There were six different presentation orders possible for the three sets of cards (e.g., ABC, BAC, CAB, etc.); each of the six presentation orders was utilized an equal number of times within the experiment.

The specific materials presented to the subjects consisted of 225 3x3 white response cards, upon which the forms had been stencil-painted. The 12 stimulus cards were placed on a white background board (18"x24"). This board was placed upright before the subject. The stimulus cards were arranged in three rows of four cards each; under each stimulus card was a narrow slot through which the subject dropped the response cards. At the top of the board was a thirteenth slot for which there was no stimulus card and which was designated "no sort" (the reason for this will be made apparent in the report of the instructions given to the subjects). On the back of the board, facing the experimenter, the
slots were labeled alphabetically A through L. The cards were so arranged that any given card could be correctly placed only in one of the thirteen slots.

Answer forms were prepared containing the card number, the alphabetical letter which designated the correct slot placement for each card, and space for recording the subject's actual response. Thus, as the subject sorted a card, the examiner gave immediate information as to the accuracy of the sorting and simultaneously recorded the response for later analysis.

The instructions given each subject

"This is a task involving concept formation. You may have heard of it in your psychology class. This was originated at Wisconsin and has been used at Iowa, Minnesota and Michigan. We have norms for the task from these universities. We are trying to explore this task further and gather data from our university. However, there is enough information to date that you will be told what your performance has been compared to that of others.

Look at this board. If you compare these 12 cards with each other you will find that they differ in several ways. (pause) There are different colors, forms, numbers and positions. (These were carefully pointed out and particular care was taken to make sure that the subject understood that "position" referred to the quadrant of the card and not to the configuration of the figures.)

Now, if you had a card like this (pointing to any of the 12 stimulus cards) you could sort it with these (demonstrating the possible sortings) because it is the same color, or the same form, or the same position, or the same number of figures. (pause) Is that clear?

The task you are to do will be somewhat more difficult. You will be given cards to sort, but not on just one of these ways; instead you will have to sort the cards on a combination of two ways. As you can see there are many possible combinations of the four dimensions. (pause) When you are given the cards your task will be to sort them on the correct combination of two of the four dimensions.

This is the way you will do it. You will take the cards one at a time in the order they are given to you, and attempt to match it with
one of the cards on the board. The matching should be on the basis that they are both alike in two ways. You place the card you are sorting through the slot under the card you think it matches. As soon as you do, you will be told whether it was correct or not. If it was a wrong sort, the buzzer will sound like this .... if it is correct, there will be no sound. (The buzzer was demonstrated several times, with repetitions of the last line of instructions.) As you can see, this task at first is a kind of trial and error until you find the right combination.

There are two things you will discover as you do this task. First, that when you find the correct combination and use it, each card can be placed correctly in only one place on the board. For example, the first card might be correctly placed here (pointing to any stimulus card) and then it would match no other card on the same combination; whereas, this second card, in the same two ways, could go only with this card (pointing) and no other, and so on.

Second, have you noticed this top slot which has no card above it? This is the "no-sort" slot and is to be used just as the others are. You will find when you are sorting on the correct combination that there will be cards in the stack that will not match any of the 12 cards on the board. Then you are to put the card in the "no-sort" slot to be correct. This is not a trick or gimmick, you can be just as right and as wrong using this slot as any of the other twelve.

Are there any questions? (Pertinent questions were answered by repeating the appropriate section of the instructions; the subject was asked to defer all other questions until the end of the experiment.)

There are three different sets of cards to be sorted. Each set has 75 cards in it. You will be given one set at a time, and you are to sort them in the order given to you. When you finish each set you will be told how you scored in comparison to others. Further, the sets differ from each other in level of difficulty. You may have the easiest first or the most difficult, but you will not know this until the task is completely finished.

You will have fifteen minutes for each set. Some finish in this time, some do not. You may use as much or as little time as you wish on any one card. Your score will be made up of the number correct, the time, and the number sorted. You will be told how you did at the end of each set."

The success and failure experiences

The failure and success reports were the same for all subjects and were administered in the manner to be described. At the end of
the presentation of a particular set of cards, the examiner supposedly scored the task, and then, after referring to fictitious norms, reported the "results" to the subject. The specific experience of failure or success was interpreted to the subject in the following manner:

**Failure** — "Well, you didn't do very well that time. Your score was 13% below the average for that set of cards. The score is based on the amount done in the time allowed and the percentage correct. On that basis you were quite a bit below the norm...just about 13% below the average for this particular set...well, let's try the next set."

**Success** — "Oh, you did very well on that set. Your score was 19% above the average for that set of cards. This set seems to be difficult for many people so that 19% above the average meant that you did quite well on it...let's try the next one now."

The rationales for some of the "information" given in the above statements was based on a number of considerations. It was possible that the subjects could have been aware of the fact that they had performed adequately in terms either of speed or accuracy. This possibility necessitated complex statements regarding the "failure". Therefore, in the failure report, the explanation of how the "score" was determined stressed both time and accuracy. In the success report, the inclusion of the statement about the difficulty of the particular set of cards was based on a similar rationale. Some subjects were scheduled to receive "success" when objectively they had either not completed the sorting and/or had very few correct responses. By stressing the difficulty of the set, it was thought that the success report might be made more plausible and therefore, more acceptable to the subject.

A final statement of performance was given at the end of the ex-
experiment and was the same for all subjects. It consisted of a general report to the effect that the subject had done quite adequately despite difficulty with one set, and that the over-all performance had been "well above average".

The inclusion of these statements in the final instructions was based upon information obtained in pre-testing, which was aimed at exploring the failure-success reports as well as testing the WCST. The judgment of the efficacy and plausibility of the reports was based on comments made by the pre-test subjects during the test and in an informal post-experimental discussion.

As would be expected, there were many spontaneous comments during the experiment, particularly after the failure report had been given. These consisted of statements such as "I'm not doing very well on this set" or "this one is harder than the other". After the experiment the experimenter made a comment to the effect that some sets had seemed easy and others difficult for the subject, and then the subject was asked what he thought about the tasks. Many of the subjects reacted with statements typified by the following: "I certainly did mess-up that one section" or "Well, I was mixed-up for awhile and then I got it".

A subject who had objectively improved his performance frequently would state spontaneously (or agree with the examiner after the reported failure) that his performance had deteriorated. Thus, the reports of "success" or "failure" seemed to have *prima facie* validity. At least, this may be inferred from the apparent absence of any relationship between the subject's objective success and/or failure and his readiness
to accept or agree with the examiner's evaluation of his performance.

Subjects

At the beginning of the summer quarter the Taylor Manifest Anxiety Scale was given by the instructors to all of the sections of the 401-402 Introductory Psychology course at The Ohio State University. The students were informed that it was part of the course requirement, thereby tending to minimize any association between the test and the subsequent experiment.

The experimenter then met with the various psychology classes and structured the selection of the subjects in the following manner: Each class was told that there was a general research project which required a random, cross-sampling of the students within the 401-402 courses. Therefore, names had been drawn from the class rolls and these individuals were requested to serve as experimental subjects. The list of subjects from the class was then read and these students filled out a schedule, indicating their available free time. From these schedules each student was assigned an individual experimental hour.

The students actually chosen as experimental subjects were those who had scored in the upper and lower twenty percent of the distribution of TMAS scores. Initially there were approximately 80 subjects. However, through attrition from pre-testing, course drop-outs and scheduling difficulties, a total of only 56 subjects was available for the final experiment.

There were 28 anxious and 28 non-anxious subjects; the mean
score of the TMAS for the anxious group was 23.76, and, the range was from 20 to 32; the mean score of the TMAS for the non-anxious group was 3.75, and, the range was from 0 to 6.

The 28 subjects in each group were placed in rank order according to their TMAS score. On the basis of alternate ranks, each was divided into two sub-groups, each having 13 subjects. After the sub-division the mean TMAS scores for the two anxious groups were 23.75 and 23.78; the mean TMAS scores for the two non-anxious groups were 3.71 and 3.78.

One sub-group of the anxious subjects was then assigned to the experiment, receiving first success, followed by failure; the remaining sub-group then received failure first, followed by success. The same assignment was made of the two sub-groups of the non-anxious group.

Inasmuch as knowledge of the subject's anxiety scale score could have influenced the experimenter's handling of the "success" and "failure" experiences, the actual assignment of subjects to the groups was done by another person. The information the experimenter had about each subject at the time of the experiment was the subject's name, a code letter indicating the order in which success and failure were to be administered, and, a code number indicating the order in which the three sets of cards were to be presented to the subject.
RESULTS

The design of the experiment was to administer a trial, present the success (or failure) experience, administer the second trial, present the failure (or success) experience, administer the third trial, and, end with a general statement of "the above-average performance" of each subject.\(^4\)

The measure of performance in this study was the percentage of correct responses for a given trial. This particular measure was utilized so that (a) any variation in the number cards actually sorted on a given trial (on either an intra- or inter-individual basis) could be taken into consideration, and, (b) the scores would be comparable.

In order to determine first, whether any differences found should be attributed to differences in initial learning (rather than to the experimental variables), and, second, whether group differences in

\(^3\)"Trial" is used in this chapter to mean the presentation of one set of 75 cards. This term is used rather than "set" as the latter term would be misleading in that any of the three sets of 75 cards could be presented to a subject as the first, the second, or, the third.

\(^4\)The following abbreviations will be used throughout the remainder of the dissertation: A, anxious; NA, non-anxious; F-S or S-F to indicate the order in which the failure and success experiences were presented to any given group.
learning occurred during the experiment, the significance of differences between the means of the percent of correct scores for each group were determined. These means of the percent of correct scores (hereafter referred to as "mean accuracy scores") for each experimental group on each of the three trials are presented in Table I.

Table I

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A:F-S</td>
<td>14</td>
<td>38.5%</td>
<td>39.7%</td>
<td>60.0%</td>
</tr>
<tr>
<td>NA:F-S</td>
<td>14</td>
<td>41.6%</td>
<td>56.5%</td>
<td>59.2%</td>
</tr>
<tr>
<td>A:S-F</td>
<td>14</td>
<td>40.1%</td>
<td>51.5%</td>
<td>59.2%</td>
</tr>
<tr>
<td>NA:S-F</td>
<td>14</td>
<td>29.7%</td>
<td>41.6%</td>
<td>60.0%</td>
</tr>
</tbody>
</table>

Tests of significance (t-tests) were performed between each set of trials for each experimental group. Between-group comparisons also were made in order to test differences between the mean accuracy scores of NA and A on a given trial under the same experimental condition (either S-F or F-S). In addition, comparisons were made which tested for differences between mean accuracy scores of all A to all NA regardless of the order of presentation of the experimental conditions. There were a total of seventeen possible combinations for which t-tests were computed — only two were or tended toward significance. These are presented in Table II.
Table II

Significant t-tests

<table>
<thead>
<tr>
<th>Group</th>
<th>Trial comparisons</th>
<th>t</th>
<th>Level of probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A:F-S</td>
<td>2-3</td>
<td>2.21</td>
<td>.05 – .02</td>
</tr>
<tr>
<td>NA:S-F</td>
<td>2-3</td>
<td>2.08</td>
<td>.10 – .05</td>
</tr>
</tbody>
</table>

Chi-square was used initially to investigate differences in performance between the groups. The statistics were set up simply on the basis of increase or decrease in performance from one trial to the next. A given individual's score was recorded as a plus or minus, depending on whether he had increased or decreased in performance from the previous trial. The numerical tally of pluses and minuses were the data utilized in the Chi-square.

Following a similar model of the t-tests, the Chi-squares were computed on an intra-, as well as, inter-group basis. There were eight possible combinations for which the Chi-squares were computed — of these, only two were significant.

Although Chi-square did yield some significant results, it was considered an inadequate statistic in that it was based solely upon increase or decrease in performance and did not or could not take into account amount of change in either direction. This could have a distorting effect in that minimal change, which could be a chance phenomenon, would be given as heavy a weighting in the total score as very marked deviations. Another statistic was sought because of this, and, because the mean accuracy scores increased with successive trials
(c.f. Table I). The Mann-Whitney U-test was the statistic finally selected as it met the double criterion of being relatively sensitive both to the amount of change and the direction.

There were twelve possible combinations of intra-group and inter-group comparisons for which U-tests were computed. The results are presented in Table III and Table IV.

Table III

U-test comparisons of differences between trials

<table>
<thead>
<tr>
<th>Group</th>
<th>Trials</th>
<th>U value* #</th>
</tr>
</thead>
<tbody>
<tr>
<td>A:F-S</td>
<td>1-2 and **2-3</td>
<td>1.84</td>
</tr>
<tr>
<td>NA:F-S</td>
<td>**1-2 and 2-3</td>
<td>.69</td>
</tr>
<tr>
<td>A:S-F</td>
<td>#1-2 and 2-3</td>
<td>.92</td>
</tr>
<tr>
<td>NA:S-F</td>
<td>1-2 and #2-3</td>
<td>1.06</td>
</tr>
</tbody>
</table>

*A U-value (for a one-tailed test) of 1.61 is necessary for significance at or above the probability level of .05; the significant U-value in Table III is underlined.

**This sign indicates the direction of differences between the sets of trials under comparison. In Table III ** indicates which of the two differences had the greater mean accuracy score gain.
Table IV

U-test comparisons of differences between groups

<table>
<thead>
<tr>
<th>Trials</th>
<th>Groups</th>
<th>U value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &amp; 2</td>
<td>A:F-S &amp; **NA:F-S</td>
<td>2.07</td>
</tr>
<tr>
<td>2 &amp; 3</td>
<td>**A:F-S &amp; NA:F-S</td>
<td>2.07</td>
</tr>
<tr>
<td>1 &amp; 2</td>
<td>**A:S-F &amp; NA:S-F</td>
<td>.83</td>
</tr>
<tr>
<td>2 &amp; 3</td>
<td>A:S-F &amp; **NA:S-F</td>
<td>1.36</td>
</tr>
<tr>
<td>1 &amp; 2</td>
<td>A:F-S &amp; **A:S-F</td>
<td>2.26</td>
</tr>
<tr>
<td>2 &amp; 3</td>
<td>**A:F-S &amp; A:S-F</td>
<td>1.68</td>
</tr>
<tr>
<td>1 &amp; 2</td>
<td>**NA:F-S &amp; NA:S-F</td>
<td>.74</td>
</tr>
<tr>
<td>2 &amp; 3</td>
<td>NA:F-S &amp; **NA:S-F</td>
<td>1.15</td>
</tr>
</tbody>
</table>

*A U-value (for a one-tailed test) of 1.64 is necessary for significance at or above the probability level of .05; the significant U-values in Table IV are underlined.

**This sign indicates the direction of differences between the groups under comparison. In Table IV ** indicates which of the two groups had the greater mean accuracy score gain.

The data in Table III indicate that there was only one significant difference among the intra-group comparisons. The A:F-S had a significant increase in accuracy score following success. There was a trend towards a significant difference in the NA:S-F group which had an increase in accuracy score following failure:

The data in Table IV indicate the following:

First, there were no significant differences between the A and the NA groups who were tested under the pattern of S-F.
Second, significant differences were found between the A and the NA groups who were tested under the pattern of F-S. These differences are as follows:

(a) the NA had a greater number of the group increasing in accuracy score following initial failure than did the A group.

(b) the A had a greater number of the group increasing in accuracy score following subsequent success than did the NA group.

(c) the A had a greater number of the group increasing in accuracy score following subsequent success than did the same A group following initial failure.

Third, the results of the comparisons of A:F-S to A:S-F and of NA:F-S to NA:S-F were:

(a) on the second trial a significantly greater number in the A:F-S group increased their accuracy scores than did those in the A:S-F group.

(b) on the third trial a significantly greater number in the A:F-S group increased their accuracy scores than did those in the A:S-F group.

(c) no significant differences were found in any comparisons of the NA:F-S to the NA:S-F.
CONCLUSIONS

Two qualifications need be made with regard to the U-test results. First, any interpretation of the results must take into consideration what appears to be a major limitation in the experiment itself. This is the limitation placed on the data by the experimental task. Examination of the accuracy scores of the final trial for all of the groups shows that there is practically no variation. This lack of variation would indicate that either a plateau or a ceiling had been imposed upon the subjects' performance by the task. While this would not affect the interpretation of results on the previous trials, it would perhaps modify (although not negate) the meaning of significant differences found on the last trial and would certainly make for questioning of the non-significant differences.

Second, the U-test results must be related to the accuracy scores for each group under each condition (Table I). These accuracy scores show whether the U-test statement of differences in score between any two groups is relative or absolute. In other words, differences between groups could be obtained in different ways; for example, a difference could occur if one group made improvement and the other declined, or if one group maintained its level of performance and the
other group changed markedly, either improving or declining. The manner in which differences occurred between groups would clearly affect the interpretation of the U-test statistics.

Thus, support and/or negation of the hypotheses will be found in both the U-test data and the mean accuracy scores which indicate the direction of change. Both sets of data indicate that although none of the hypotheses was supported in its entirety, there was partial affirmation of some of them. A restatement of the hypotheses and the various results is as follows:

(a) Hypothesis: the A:F-S group will decrease in performance after failure and will increase performance after success.
Result: the A:F-S group maintained its level of performance after failure and increased its performance significantly after success.

(b) Hypothesis: the NA:F-S group will increase performance after failure and will decrease performance after success.
Result: the NA:F-S group increased performance significantly after failure and increased performance, although not significantly, after success.

(c) Hypothesis: the A:S-F group will increase performance after success and will decrease performance after failure.
Result: the A:S-F group increased in performance, although not significantly, after both success and failure.

(d) Hypothesis: the NA:S-F group will decrease in performance after success and will increase performance after failure.
Results: the NA:S-F group increased in performance, although not significantly, after both success and failure.

It is apparent from the data presented in a variety of statistical forms that the specific conclusions which can be reasonably drawn from this research are extremely limited.

First, the statement can be made that success and failure do not appear to have the same effect upon performance. Success automatically seems to engender further success; whereas, failure's effect upon performance appears to be a function not only of the given affective state of the individual but also of the immediately preceding reinforcement experience.

Second, that the differences in affective state of the anxious and non-anxious do result in a differential reaction to failure. However, this differential reaction can be avoided through previous experience or eliminated by subsequent experience.

Third, the interaction of the two variables, along with the sequence effect of the reinforcement seem to combine in such a manner as to modify the "isolated" functioning of each an extent great enough to alter the predicted form of reaction.

It is clear that the above mentioned conclusions cannot be explained in terms of simple change in general drive state, no matter how tortuously this concept is manipulated; nor is it possible to account for the results with any modification of theory such as was suggested by Lucas in his statement "an obvious assumption is that responses produced by anxiety and failure increase at a rate different
than those produced by corresponding drives" (28, p.65).

The necessary question then becomes "why do the results not follow the predictions or fit within a theory?" Part of the answer comes from examination of the research studies that did obtain predicted results. These studies had in common one or all of the following points: (1) the experimental task was of a "simple" nature, not particularly related to daily experience; (2) when positive and/or negative reinforcement was used it was usually of a "contrived" nature (such as shock-no shock) and further, was not used in sequence; (3) "anxiety" was often the only experimental variable under consideration.

Another part of the answer comes from an examination of the studies that did not find readily explicable results. These studies, for the most part, were those which attempted (1) to handle personality and learning variables together and (2) used more realistic experimental situations and forms of reinforcement.

These two different sets of results could actually be anticipated if the first set is viewed as resulting from studies aimed at what might be called "basic laboratory research" and the second set viewed as stemming from efforts to extend this basic research into problems met with in the daily environment. The conclusion from these two sets of data would seem to indicate that variables dealt with in isolation do not always react in the same fashion when interaction with other variables occurs. And, further, that some "basic research" may have little functional meaning.

This does not imply that "anxiety" is a useless concept; but
rather implies that its original formulation, as a simple motivation-al factor contributing to general drive state, is far too limited to explain the various data.

This does not imply that "success-failure" is a useless concept; but rather implies that formulating it simply as opposite sides of the same coin, (i.e., that the two factors have equal weighting whether effecting habit strength or altering the affect of the individual) is again far too limited to explain the various data.

What this does imply is that the results of many experiments would be a function of two dimensions, usually quite separate and apart from the discrete experimental variables under consideration: the first is the interaction effect of the experimental variables; the second is the meaningfulness of the experimental task and experimental conditions to the subjects.

While granting the need for "basic research", the question can be raised as to whether there is any human learning that is not influenced by the individual's personality and by the situation in which he finds himself. And, the question must also be raised as to whether "personality" can be studied without having, at least implicit, the question "and what effect does this have upon performance?" It would seem that the confusion of the results of this study would add, rather than subtract, weight to the need for more consistent research which utilizes both personality and learning variables.
The statistical data and the conclusions drawn from these data, as presented in the two final chapters, followed logically from the formulation of the original hypotheses. These hypotheses were stated in absolute terms, i.e., that there would be a definite increment or decrement in performance, the occurrence of either being a function of the combination of a specific subject-group with a specific experimental condition.

It was apparent from an inspection of the data that factors other than the experimental variables effected the performance of the subjects. One of these factors, as has been previously discussed, was the experimental task itself. The task appeared to allow for improvement in performance with successive trials; this improvement was, for the most part, independent of the effect upon performance of the experimental variables under consideration. This improvement with practice would weight strongly against any prediction of a decrement in performance and strongly for any prediction of an increment in performance.

It would seem logical to inquire, although clearly on a post-hoc basis, as to whether the data would have been different or presented a different pattern if practice or learning effects had been controlled.
One method through which the effects of learning on the obtained results could be, in part, controlled or partialed-out would be by reformulating the hypotheses in terms of relative amounts of increment in performance rather than in terms of absolute increment or decrement in performance. In other words, the assumption would be that learning would occur with all subjects when given a series of trials, and, thus, the effects of the experimental variables would be to differentially decrease or increase the amount of learning, or, restated, make for different amounts of increment.

Utilizing this "new" assumption, a necessary restatement of the original hypotheses would then be (a) failure would result in less of an increment in performance than would success for anxious subjects, and, (b) success would result in less of an increment in performance than failure for non-anxious subjects.

To test this general hypothesis a comparison may be made between two groups -- a group of predicted "lesser" increment in performance and a group of predicted "greater" increment in performance. The former group would contain the discrepancy scores of the anxious subjects following failure and the non-anxious following success; the latter group would be composed of the discrepancy scores of the anxious subjects following success and the non-anxious following failure. A model of this re-organization of the original experimental groups is presented in Table V.
A t-test of the significance of the differences between the mean discrepancy scores for Group A and Group B was performed. This difference was significant at the .05 - .02 level (t 2.22).

The interpretation of this significant difference is difficult. This is that although the revised grouping of the subjects does tend to control practice or learning effects (through the equal placement of early and late trials in both groups), it also serves to eliminate the sequence effect of success-failure or failure-success upon performance. Essentially the elimination of the sequence of success-failure alters the original experimental design so that the effects of failure are dealt with quite independently of the effects of success upon the performance of anxious or non-anxious subjects.

Thus, it would appear that this supplementary analysis of the obtained data (a) does not seem to offer major contradictions to the conclusions stated in the final chapter, but, (b) does seem to offer
further confirmatory evidence that success or failure, anxiety or non-anxiety are variables which do have an independent effect upon an individual's performance.


42. Rosensweig, S. Further comparative data on repetition-choice after success and failure as related to frustration tolerance. J. genet. Psychol., 1945, 66, 75-82.


51. Spence, K. W., & Taylor, J. A. Anxiety and the strength of UCS as determiners of the amount of eyelid conditioning. J. exp. Psychol., 1951, 42, 183-188.


I, Jean Lucille Burton, was born in Cleveland, Ohio, March 13, 1925. I received my secondary school education in the public schools of Cleveland and Lakewood, Ohio. My undergraduate training was obtained at Ohio State University, from which I received the degree Bachelor of Arts in 1949. The academic year 1949-50 was spent in graduate work in psychology at the Merrill-Palmer School in Detroit, Michigan. From the Ohio State University, I received the degree Master of Arts in 1951. While in residence at the Ohio State University, I was assistant to Dr. George A. Kelly during the year 1950-51, held a teaching assistantship from June, 1951, through June 1952, and was an assistant instructor in the Department of Psychology during the year 1953-54. In the year 1952-53 I received a United States Public Health Scholarship for the clinical internship in the Dearborn Public School System, Dearborn, Michigan. During 1954-55 I held the position of clinical psychologist at the Psychological Clinic, Rutgers University. In 1955 I accepted my current position of Instructor in the Department of Psychology, Douglass College, Rutgers University.