AMRL-TR-65-39

MOTIVATIONAL CORRELATES OF INDIVIDUAL DIFFERENCES IN PERFORMANCE

MARK W. STEPHENS, PhD
K. M. MICHELS

PURDUE UNIVERSITY

MAY 1965

BEHAVIORAL SCIENCES LABORATORY
AEROSPACE MEDICAL RESEARCH LABORATORIES
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Requests for copies of this report should be directed to either of the addressees listed below, as applicable:

Federal Government agencies and their contractors registered with Defense Documentation Center (DDC):

DDC
Cameron Station
Alexandria, Virginia 22314

Non-DDC users (stock quantities are available for sale from):

Chief, Input Section
Clearinghouse for Federal Scientific & Technical Information (CFSTI)
Sills Building
5285 Port Royal Road
Springfield, Virginia 22151

Change of Address

Organizations and individuals receiving reports via the Aerospace Medical Research Laboratories automatic mailing lists should submit the addressograph plate stamp on the report envelope or refer to the code number when corresponding about change of address or cancellation.

Do not return this copy. Retain or destroy.

800 - July 1965 - 448-49-1049

MOTIVATIONAL CORRELATES OF INDIVIDUAL DIFFERENCES IN PERFORMANCE

MARK W. STEPHENS, PhD
K. M. MICHELS

FOREWORD

This research was initiated by the Behavioral Sciences Laboratory of the Aerospace Medical Research Laboratories, Aerospace Medical Division, Wright-Patterson Air Force Base, Ohio.

The research was conducted by Dr. Mark W. Stephens and K. M. Michels of Purdue University under Contract AF 33 (616)-7962 with the Training Research Division, Behavioral Sciences Laboratory, Aerospace Medical Research Laboratories. The research was performed in support of Project 7183, "Psychological Research on Human Performance," Task 718305 "Behavioral Effects of Environmental Stress." W. Dean Chiles, PhD, of the Training Research Division served as the contract Monitor for the Aerospace Medical Research Laboratories.

The report covers research accomplished between May 1961 and January 1965.

This technical report has been reviewed and is approved.

WALTER F. GRETHER, PhD
Technical Director
Behavioral Sciences Laboratory

ABSTRACT

A three-year program of research was directed at the development of "paper and pencil" measurement techniques that would permit the assessment of the potential "motivatability" of subjects in experiments concerning the effects of environmental stress on human performance. A largely empirical approach was used in this research. Performance measures of a large number of subjects on several different tasks were used as the criterion measures in item analyses of several personality inventory tests. The resultant pool of cross-validated items will, it is hoped, represent a step toward increasing the precision of performance research.

PREFACE

The control of individual differences has been a persistent problem in conducting research on the performance effects of environmental stresses. In carrying out such research, it would not be sufficient to eliminate individual differences through selection procedures even if that were feasible. In attempting to apply the resultant data to real world, practical situations, it is a fact of life that people do differ. And, it may well be that the most important contribution the behavioral scientist could make is the specification of the interactions of these individual differences with environmental parameters. More specifically, when the man is exposed to environmental stresses that are less than overwhelming but still at levels that would reasonably be expected to have an effect on his performance, an altogether too frequent result is a marked redistribution of performance. Putting it differently, perhaps more so in stress experiments than in other psychological research areas, subject differences and their interactions with the experimental conditions constitute an undesirably (sometimes unacceptably) large portion of the total variance. The result is that the effects of the primary variables are obscured.

A variety of theoretical, experimental and anecdotal observations suggest that the solution of this problem lies in the specification of the level of motivation of the subject in interaction with stress level since the effect of stress is in fact, for the most part, an interaction between stress level and motivation. To the extent that this is true, proportionate increases in the precision of the results of stress research should be achieved if techniques can be developed for reliably assessing motivatability.

The ultimate goal of this program as defined by the Behavioral Sciences Laboratory was to improve stress research methodology. Thus, in an important sense, the success or failure of this effort cannot be properly judged except on the basis of the successes and failures of the resultant techniques when they are applied to specific stress research problems.

W. Dean Chiles
Contract Monitor

TABLE OF CONTENTS

LIST OF TABLES

SECTION I
INTRODUCTION

Research concerning the effects of environmental conditions on performance is plagued by the problem of individual differences. Different subjects behave differently in the same conditions, and react differently to the same changes in conditions. As a result, greater numbers of subjects are required to achieve stable results, often at considerable cost; at times, some otherwise important investigations may even become practically impossible. Furthermore, these differences represent limitations in the extent to which conclusions can be generalized across subjects.

If these individual differences could be accounted for and statistically controlled, an important methodological obstacle could be removed from laboratory studies. The purpose of this research was to attempt to make a first step toward resolving this problem.

The specific goal was to devise a test, or battery of tests, which might predict individual differences in laboratory task performance. The particular interest was in tasks (eg, classical conditioning, verbal learning) which are often employed in studies of the effects of environmental stress. More particularly, the purpose was to develop a test of motivational differences related to performance, since it is commonly assumed that motivation is largely responsible for differences in performance of simple tasks.

At face value such an undertaking appears simple and straightforward. In execution, however, it entailed unusually complex methodological and conceptual problems. For example, there is wide disagreement as to the nature of "motivation." It was necessary, then, to review the various theoretical definitions of motivation, before articulating the variables which should be measured. It was expected in advance that these variables would have different effects in different conditions; this, then, required systematically varying conditions while collecting criterion data to validate tests. Several other problems complicated the research.

This report covers three separate general topics. First discussed (sections II and III) are the theoretical and methodological problems which dictated the strategy of the research. This discussion is included for the reader who is interested in the broader theoretical implications of the research, or who is engaged in similar research. Next, the results of the twelve separate studies conducted are reported. These included five studies employing reaction-time tasks (Section IV), three concerning finger-retraction conditioning (Section V), two regarding verbal paired-associates learning (Section VI), and two with a complex verbal learning task (Section VII). The criterion data for item-analysis and cross-validation of the tests were collected in these studies. In addition, the studies were intended to clarify the nature and effect of the motivational variables measured by the tests; to clarify the effects of environmental variables of motivational nature; and to clarify the complex interactions of motivational and other variables in determining task performance. The results of these studies will be of interest to the reader interested in the nature and effects of motivational variables; they also have ramifications for the conditions in which the tests developed may be useful to other investigators. Finally (Section VIII), the tests themselves are reported, along with the procedures employed in test development, the results and implications of cross-validations, and a general summary of the results and implications of the research.

SECTION II
CONCEPTUAL PROBLEMS

DEFINITIONS OF MOTIVATION

In different theories the term "motivation" is assigned to very different classes of variables. In this research it was desirable to include as many different test-measured variables as possible, to provide a maximally useful test battery. However, it was also desirable to define these variables as clearly and systematically as possible, to provide some broader basis for generalizing results. The first step, then, was to review the various theories of motivation to (1) derive a list of variables to investigate and (2) attempt to define the relationships of these variables to one another.

At least six different theoretical approaches to motivation can be distinguished. Of these, only three were of use in suggesting variables which might be measured by tests. All six, however, are briefly discussed below, to provide the broadest possible context for later discussion.

The Classical Personality Theorists: Master Motives

The concept of motivation was introduced to psychology more by Freud than by any other man. For Freud, motivation was considered some force (psychic energy) which both energized and directed all behavior, both overt and "intrapsychic." Freud proposed that the goal of all behavior is gratification, and that gratification results from the successful diminution of psychic energy. This hedonistic view is not at all foreign to most later (and contemporary) views of motivation. In addition, the equation of pleasure with reduction of stimulation (psychic energy) is roughly equivalent to the "drive reduction" definition of reinforcement.

Freud, however, was more specific. Pleasure and motivation were seen to be primarily sexual in nature. Sex was defined broadly, and not at all clearly. Nevertheless, even though "sex" seemed often to mean any and all pleasure, Freud insisted that all pleasure is in turn, sexual. Freud continually altered his theory; and aggression was sometimes added to sex as a second primary motive, or as a primary corollary of sexual motivation. In any case, the distinctive aspect of Freud's motivation theory was its "Master Motive" character: the isolation of a particular goal or motive which was to account for all behavior.

Personality theorists following Freud accepted the Master Motive concept; they differed only as to what constituted the Master Motive. For Jung it was "self-actualization" instead of sex; for Adler, "overcoming"; for Horney and Fromm, reduction of "basic anxiety," although Fromm and Horney disagreed as to the nature of basic anxiety. The hallmark of classical personality theories is the Master Motive concept; the primary difference among them is the specification of what the Master Motive is. (Several other views of motivation, coming from sources other than the classical personality theorists, are logically equivalent to such Master Motive concepts, eg, White's "competence" and "effectance" motivation, the functionalist view of all behavior as "motivated" to enhance survival, and Woodworth's "behavior primacy.")

This approach to motivation has not contributed to the isolation of motivational correlates of individual differences in performance; it has not because logically it cannot. Such an approach is not intended for such a task. The Master Motives were proposed as psychological universals, not as psychological variables; they are consequently not measurable, since their measure is all behavior; and they are without differential effects on behavior, since they are responsible for all behavior. Individual differences in behavior can be accounted for only in terms of different manifestions of the same basic Master Motive. Master Motives are, then, more metatheoretical than theoretical concepts.

These conceptions of motivation have contributed immensely to social philosophy, clinical practice, and even to suggestions for research (eg, research concerning personality development as "psychosexual" development). Nevertheless, they are logically not suited to contributing to the search for individual differences in motivation, although they are the conceptions of motivation held by many psychologists.

Homeostasis

The first systematic use of the concept "motivation" in experimental psychology came through the work of Cannon (ref. 1), following a concept previously proposed by the physiologist Bernard (ref. 2). Any disruption of the equilibrium of the "milieu interior," or departure from the ideal physiological and biochemical state of the organism, threatens survival. To survive, the organism must have mechanisms for readjusting the balance. Many internal biological mechanisms serve such a function. Beyond these mechanisms, however, Cannon (and, by implication, Bernard) proposed that the organism's behavior was such a mechanism. Behavior, then, was seen as being directed wholly toward maintaining homeostasis. Consequently, the physiological states (disequilibria) which comprised "motives" were responsible again for both the mobilizing and the directing of the organism's behavior. In contrast to Master Motive conceptions, however, homeostasis concepts provided for many motivational variables, and consequently for some differential prediction of behavior.

This view dominated the research concerning, and even the textbook treatment of, motivation until little more than a decade ago. Cannon and many others demonstrated that behavior could be predictably manipulated by manipulating the homeostatic conditions, or biological needs, of the organism. It was repeatedly demonstrated (although recently refuted by Campbell and Sheffield, ref. 3) that the sheer presence and extent of activity was a function of the existence and extent of one or more biological need states.

It was also demonstrated that "motives" or "needs" could be learned, ie, that the organism could be shaped to respond to states and substances irrelevant to survival in the same way it responded to survival-relevant states. These "secondary needs" or "secondary drives" were then proposed to account for behavior not attributable directly to biological needs. They attained their importance, however, from their association with the primary biological needs, which therefore still remained the "basic" source of motivation and behavior.

Homeostatic theory, at least in its universal form, has undergone considerable decline. On closer view, there are too many biological needs which do not affect behavior, and too many states which do affect behavior but are not directly survival-relevant, to permit acceptance of homeostasis as the general basis of motivation and behavior (ref. 4). The discovery of spontaneous activity in neural tissue made it unnecessary to postulate some state or force to account for all neural, and organismic, activity. The "environmentally induced drive" research of Harlow and others, and subsequent research concerning "novelty," "curiosity," "alternation behavior," and "exploratory behavior," reduced the apparent power of homeostatic theory, even though such research can be encompassed by it. "Sensory deprivation" research embarrassed homeostatic theory as well as other approaches to motivation. But most particularly, homeostatic theory could not overcome two major theoretical limitations. First, to demonstrate that motives can be learned from primary needs does not help determine what motives are learned, or how they can be measured, or how they are aroused; consequently, homeostatic theory leaves a great deal unanswered. Second, it became more and more apparent that to attribute both the mobilization and the direction of behavior to the same states and/or variables was oversimplification.

Homeostatic theory, then, was also of limited use in directing a search for tests of motivation. It cannot, however, be totally dismissed. Pribram (ref. 5), for example, has revived and revitalized many homeostatic concepts with his conception of the "biased homeostat" as a model to encompass considerable neuropsychological data. And the "need arousal" model for construing motives, ie, the view of motives as states which are aroused by internal and/or external conditions, persists in many of the contemporary molar approaches to motivation, eg, McClelland (ref. 6) and Schachter (ref. 7).

Hedonic Theory

The hedonic conception of motivation is the oldest (philosophically, at least), the most general, but also the least important in this discussion. Simply, hedonic theory proposes that the organism behaves as it does as a function of seeking pleasure (and, generally, avoiding pain). The view is as old as philosophical hedonism. In psychology it dates to Thorndike's Law of Effect. It corresponds most closely with popular notions about "motivation." It more or less describes Freud's view of motivation, and all other Master Motive conceptions, if not, in fact, all views of motivation.

Hedonic theories can, however, be distinguished from others. In Freud, the specific source of all behavior was psychic energy. Pleasure might be generally associated with discharge of such energy, but at best as an epiphenomenon: the primary phenomenon was the accumulation and discharge of the energy itself. Similarly, pleasure and pain might be associated with the going and coming of biological need states; but in homeostatic theories it is these states, not the associated hedonic experiences, which constitute the source of behavior. The distinctive characteristic of hedonic theory, then, is that it is the subjective state, the experience of pain and pleasure, which is sought out as the basis of motivation. Therein lies the problem of employing hedonic conceptions in experimental research: the variable itself, pleasure, is, since subjectively defined, unobservable.

Nevertheless, at least two bodies of research have grown from hedonic theory: P. T. Young's experimental hedonism (ref. 4) and the more recent neuropsychological research of Olds (ref. 8) and others. Young experimented extensively on stimulus preferences (especially food preferences) in the context of hedonic theory. Like the proponents of the Law of Effect as the cardinal principle of motivation (or of reinforcement), however, these efforts cannot escape the circularity of having to define and measure preferences (pleasure, hedonic values) by means of the same behaviors these preferences are to explain. Olds' work partially obviates this circularity. Pleasure and pain are manipulated by directly stimulating various subcortical centers. The hedonic state is defined, then, without direct reference to the behavior it is to explain. As Miller (ref. 9) and others have pointed out, however, there are many ambiguities and inconsistencies in the data supporting neurological hedonic theory. Also, the location of pleasure and pain centers still requires a behavioral criterion of apparent pleasure and pain, so the circular definition of hedonic states is not completely avoided.

In any case, the hedonic theories again are of limited use in the present context, primarily because they do not permit independent measures of motivational differences which might be related to performance.

The three remaining approaches to motivation suggest different but feasible approaches to the task at hand.

Drive Theory and "Drives" Theory

In homeostatic theory, the biological states responsible for behavior were traditionally referred to as drives, eg, sex drive, hunger drive, etc. Learned motivational tendencies were denoted "secondary drives," eg, the drive for power, the drive for success, etc. In each case it was assumed that (1) the drive represented some state of the organism which was (2) aroused by internal and/or external stimuli or events and which (3) mobilized and (4) directed behavior. The "Drives" theory was, then, homeostatic theory.

However, the term "drive" was frequently used quite loosely. Many efforts were made to compose a definitive list of the primary drives, and these lists often included states quite remote from any clear biological condition representing a direct threat to survival (eg, sex drive). At the extreme, various authors attempted first to observe behavior, then to inductively arrive at the various basic goals toward which it was directed, and then to propose that each such goal was represented by some otherwise undefined drive. Many such efforts ended in a close approximation to MacDougall's classification of the instincts. Such efforts could not transcend their own circularity, measuring the extent of the drive from observation of the behavior it was proposed to explain. As useful explanatory concepts, then, these drives met the same fate as MacDougall's instincts.

From this variant of hoemostatic theory, however, there developed contemporary Drive theory, as proposed by Hull (ref. 10) and recently extended and clarified by Spence (ref. 11), Brown (ref. 12) and others. The primary distinction between "Drive" and "Drives" theory is not just the singularization of the term, but the definition of Drive as a state which only mobilizes behavior. The directionality, or steering, of behavior is attributed to other variables. There may be many organismic states (as well as external stimulation) which are sources of Drive, or contribute to momentary Drive level. But these states (hunger, thirst, etc) are not "drives" themselves. Such states in fact serve dual functions: they are sources of Drive and sources of interoceptive stimuli. Drive functions and stimulus functions of such states are separated, and the directionality of behavior, to the extent that it relates to such a state, is attributed to the stimulus function.

Spence, Brown, and other Drive theorists have not attempted to isolate the specific biological state, or mechanism, or process to which Drive refers. Spence has suggested (ref. 13) that some mechanism of the autonomic nervous system may be suspected to be the neurophysiological referent, but Drive research and theory has generally employed the concept simply as an intervening variable, and has been concerned primarily with the use of the concept in S-R theory and research.

Although Drive is assigned only a mobilizing, not a directing, function, Drive theory and research do clearly suggest an approach toward isolating some of the motivational correlates of individual differences in performance. It was proposed initially by Taylor (ref. 14), if only by implication, and later more systematically by Spence (ref. 13) that there are relatively stable differences among human organisms in Drive level. The Manifest Anxiety Scale (MAS) was constructed specifically for measuring these individual differences. Drive theory proposes that Drive (D) and Habit Strength (H) multiplicatively determine Reaction Potential (E): that is, $E = H \times D$. The organism's response depends first on the momentary stimulus complex, and then on the organism's habit hierarchy vis-a-vis this stimulus complex. To a given stimulus complex, the organism will ordinarily have several responses which are of greater than .00 probability of occurrence. On repeated trials, or confrontation with that stimulus complex, these responses will occur with frequencies related to their probabilities of occurrence, as denoted by H. As Drive increases, however, the absolute strengths of these response tendencies (E's) increase by multiplication. Since D energizes or multiplies all response tendencies equally, the absolute differences among response tendencies will therefore increase. Consequently, the function of Drive increase (whether experimentally manipulated, per Drive stimuli, or inferred from individual differences measures) is to increase the probability of that response which is dominant in the habit hierarchy vis-a-vis that stimulus complex.

This, then, suggests that performance will be directly related to Drive level whenever the correct response is dominant in the habit hierarchy vis-a-vis that stimulus complex, but inversely related to Drive when the dominant response is any incorrect response.

This pair of predictions has been the focus of most Drive theory research. In general (see ref. 15), most data confirm the prediction that Drive increase improves performance on simple tasks (especially "classical-defense" conditioning tasks, and most particularly eyelid conditioning). However, the results have been somewhat less clear and consistent in research employing individual differences measures of Drive then when Drive is experimentally manipulated. Also, the evidence is even more inconsistent and inconclusive regarding the hypothesis that Drive increase inhibits performance in complex tasks and/or other tasks in which the correct response can be assumed to be nondominant.

In Drive theory Habit variables might also be construed as having some of the properties often attributed by others to motivation. To the extent that motivation is construed as relevent to directionality of behavior, or as determining patterns of behavior oriented to particular goal-objects or goal-states, it is the Habit variables which are motivational in nature. Such motives as need for achievement, need for approval, etc, are definable in Hullian theory as higher-order habits: highly generalized, learned tendencies to emit a given class of responses (eg, achievement-seeking, approval-seeking) in a highly generalized family of stimulus situations (those with achievement cues, etc).

From Hullian theory, then, the search for motivational correlates of individual differences in performance is directed in two channels. First, measures of individual differences in Drive level must be sought (and refined), which should refine prediction of performance to the extent that the dominance of the correct response in the task situation can be identified. Second, there must be a search for higher-order habit types of motivational variables and tests to assess differential probabilites among subjects of goal-oriented responses which will enhance or inhibit performance. The search for the latter variables and tests must be largely inductive - or, alternatively, blindly empirical. Some orientation to this research will be discussed below in reference to the "goal-directed behavior" views of motivation.

Arousal Theory

Several theories have been proposed concerning "arousal" or "activation." The views have all stemmed more or less directly from neurophysiological research concerning the Ascending Reticular Arousal System (ARAS). Aside from those whose interests are primarily in the neurophysiological phenomena per se, Hebb's position (ref. 16) has been the most central.

Arousal theories, eg, Malmo (ref. 17) and Duffy (ref. 18), are very similar to the Hull-Spence Drive system. With a few exceptions, Arousal and Drive appear to be almost interchangeable, at least in stimulus and response referents which define the terms. It is noteworthy, however, that Arousal theory emerged from neurophysiological research, apparently almost completely independent of Hullian theory and the behavioral research on which it is based. Arousal theory is not as extensively articulated and systematized as Hullian theory. In view of the very close parallels between the two approaches, however, there appears to be an encouraging possibility of coordinating the data on which the two systems are based, employing each as a supplement to the other.

The two approaches diverge in two major respects. First, Arousal theorists specifically denote the ARAS as the neurological locus or mechanism involved in mediating arousal phenomena, and are particularly interested in neurophysiological correlates of arousal. As mentioned previously, Drive theorists have been less concerned with neurophysiological correlates of Drive, and have proposed (ref. 13) some autonomic nervous system mechanism as the probable mediator of Drive. Second, Arousal theorists, in postulating the relationship of performance to Arousal, have followed the Yerkes-Dodson law (ref. 19). In brief, it proposes that (1) performance increases as a direct function of Arousal ("motivation," in the original Yerkes-Dodson law) up to some point of optimum Arousal level, beyond which further increase in Arousal leads to deterioration of performance, and (2) the "optimum" level of Arousal varies inversely with the complexity of the task, being highest for the simplest tasks. In contrast, Drive theory proposes that Drive increase leads to increase in performance in any task in which the correct response is dominant, but to response decrement in any task in which some one or more incorrect responses are dominant; and, in each instance, the relation of Drive to performance is essentially linear.

The latter discrepancy between Drive and Arousal theories is of direct importance to the present research program. Both the Drive theory and the Arousal theory hypotheses regarding this relationship have been supported by considerable data, although each is confronted by contrary data. Broen and Storms (ref. 20) suggested a simple amendment to Spence's Drive theory formulation by which the latter could accommodate the phenomena predicted by Arousal theory, and Elias (ref. 15) found evidence supporting Broen and Storms' proposition. Nevertheless, the relationship of Drive, or Arousal, to performance must remain an open question.

In summary, Arousal theory and the research on which it is based reinforces the implications of Drive theory regarding the search for motivational correlates of individual differences in performance. A major dimension to be explored in each case is that of individual differences in generalized Drive (or Arousal) level. There remain conflicting hypotheses regarding the relationship of such a variable to performance (linear or curvilinear) and the task parameters on which the relationship depends (task complexity or response dominance). These questions must be considered in designing tasks against which to validate individual differences measures of Drive or Arousal.

Finally, at least by implication (in referring to "cue function" as distinct from "arousal function"), Hebb (ref. 16) also proposes that some non-arousal variables may also be found with some of the properties (specifically, the steering properties) of traditionally-defined motivational variables. These other variables would presumably correspond to the higher-order habit variables in Hullian theory. Again, however, Arousal theorists have made no specific suggestions of particular cue-function variables to explore as correlates of individual differences in performance. The suggestions for these variables can be derived only from those whose interests in motivation have been directed at goal-directed behavior.

Goal-Directed-Behavior Views

Murray

As noted previously several other approaches to motivation correspond to the cue function of Hebb and/or the Habit variable of Hull, rather than Arousal functions or Drive. These approaches might be called "goal-directed behavior" conceptions of motivation. They share these characteristics: (1) the primary interest is in describing and/or explaining patterns of overt behavior; (2) these patterns are distinguished on the basis of apparent or inferrable goals toward which separate actions are directed; and (3) individual differences in overt behavior have been of paramount concern.

Murray (ref. 21) can be taken as the progenitor of current interest in motivational phenomena of this nature. He altered the Master Motive concept of Freud and other classical personality theorists by proposing that human organisms behave as a function of many motives or needs, rather than various ramifications of the same basic motive. Murray in fact proposed a list of some 20 such needs: need for achievement (n Ach), need for deference (nDef), need for affiliation (nAff), etc.

Murray further proposed that these needs interact with environmental "presses" in determining behavior, thus moving the personality theorists' concern farther from the "intrapsychic," instinctive phenomena toward incorporation of the environmentalism of experimental psychology. He also devised the TAT technique to measure the relative strengths of these various needs in different subjects.

Despite his mention of the role of environmental presses, Murray remained mostly interested in these needs themselves as internal, phenomenal (although not necessarily conscious) states of the organism. He was not greatly interested in overt behavior, or in measurement of needs as a means of predicting overt behavior. His interest, then, was still largely in "intrapsychic" phenomena.

McClelland and others

McClelland (ref. 6) has extended Murray's conceptions. Attempting to integrate Murray's scheme with behavior theory (especially Hull), McClelland suggested a simple model in which needs are construed as intervening variables which interact with Habit variables in determining overt behavior. McClelland also adapted the TAT technique to provide a more reliable (and quantified) measure of individual differences in the relative strengths of various needs. With this measure, he and his colleagues have conducted considerable laboratory research.

In his research, McClelland, like Murray, has been less interested in the relationship of these needs to overt behavior than in these needs per se. Among his major concerns have been the environmental variables which arouse these needs and the child-rearing antecedents of individual differences in the needs. There has been some work, however, concerning the relationship of such needs to overt behavior, especially concerning individual differences in need for achievement (nAch) as related to performance of laboratory tasks and/or to academic achievement. The results of these studies have generally, although not uniformly, lent support to the Murray-McClelland scheme as a means of predicting individual differences in task performance.

Several others have contributed to the Murray-McClelland approach to motivation. Edwards (ref. 22) devised a forced-choice inventory to measure the relative strength of 15 of Murray's needs, and French (ref. 23) devised a sentence completion measure of nAch. Atkinson (ref. 24) in particular has contributed considerable research, especially concerning nAch. He has also revised McClelland's model by the addition of an "expectation" variable in mediating overt behavior, and by distinguishing more extensively between "hope for success" and "fear of failure" as aspects of nAch.

The research which has accumulated is encouraging, but it also has suggested several problems. Perhaps more important are the problems of definition and articulation of variables, to be discussed below. Also, there are problems of measurement. The TAT technique, while apparently of some validity, is extremely time consuming. Both administration and scoring require too much time to easily allow its use with large numbers of subjects. French's sentence completion technique is only somewhat less tedious. Both techniques appear to lack discriminative power. Although each shows indications of some valid discrimination of variance in the variables it purports to measure, each obviously includes considerable error. This may be an inescapable limitation of any technique whereby needs are inferred solely from stimulus-produced fantasy behavior. The Edwards inventory, while avoiding these problems, has produced inconsistent evidence of validity and little evidence of power. Social desirability and other response-set variables which distort the subject's direct self-reports may provide a largely inescapable problem in this approach to measurement.

Schachter

A somewhat related, although more restricted, approach has been that of Schachter's (ref. 7) work concerning need for affiliation. Schachter relates his rationale to Festinger's propositions regarding self-appraisal, rather than to Murray and McClelland. Nevertheless, like McClelland he construes needs as internal (phenomenal, if not conscious) states of the organism, aroused (primarily) by interpersonal environmental conditions. He is much interested in individual differences in reactivity to these need-arousing cues. His research has involved ingenious use of experimental technique, which has been notably absent in much personality research. He has not, however, contributed to solution of the problems of measuring individual differences in such needs.

Rotter

A more distinct approach to molar motivational phenomena, and the one adopted for this research project, is Rotter's (ref. 25) Social Learning Theory. This theory purports to be not specifically a theory of motivation, but a theory of molar behavior, especially interpersonal behavior, with specific interest in individual differences and approaches to behavioral change. Rotter incorporates some aspects of Murray's and McClelland's conceptions with concepts and assumptions of Hull, Tolman, and Lewin. Most important, he integrates these within a relatively articulate quasi-mathematical model. In the process, he discriminates several separate parameters of motives or higher-order habits. It is the last-named accomplishment which makes his model maximally useful in the search for correlates of individual differences in performance.

The dependent variable in Rotter's model is Behavior Potential (BP): the probability of occurrence of a given behavior, or class of behaviors, in a given environmental situation or class of situations. BP corresponds closely to Hull's E, effective reaction potential, particularly in reference to higher-order habits. The primary classes of intervening variables which determine BP are (1) the subject's Expectancies (E's) as to the probable consequences of that behavior and (2) the Reinforcement Value (RV) for him of such consequences. Rotter's E and RV correspond closely to the terms "subjective probability and "subjective utility" employed more recently in decision theory. Expectancy for any given consequence, as a probability, varies from .00 to 1.00. Reinforcement Value, however, can be measured only in a relative, not an absolute sense, ie, only ordinal scaling is possible. Because of this, the mathematical basis of the model must remain relatively gross. In summary, the probability of occurrence of a given behavior varies directly with the favorability of the expected consequence of the behavior and with the probability (subjective) of occurrence of that consequence: BP = f(E & RV).

The possibility of articulating Rotter's model with Hull's lies in two assumptions of Rotter. First, both Expectancies and Reinforcement Values are the product of learning. They therefore can be defined as different parameters of Hull's Habit variable and, more germane to the present work, as different parameters of higher-order habits. Rotter makes liberal use of stimulus and response generalization research, although he adapts these concepts in various ways. Whereas H is solely a function of N (number of prior reinforcements) in Hull, E (and to some extent RV) is a function of the proportion of prior reinforcements to nonreinforcements in Rotter. Since Rotter's concern is with the "higher-order habits" of Hull, however, this difference is not necessarily a source of incompatibility.

The second link between Rotter and Hull has to do with the role of the stimulus. Rotter, operating on a more molar level and with deference to Lewin, refers to the "psychological situation" rather than to discrete stimuli. Nevertheless, environmental conditions are relevant not only through their prior reinforcing role but also as immediate determinants of behavior. Specifically, Rotter proposes that both Expectancies and Reinforcement Values are specific to the psychological situation or to classes of psychological situations. Environmental cues do not, then, "arouse" needs, as in Murray, but are simply the contingencies with which the needs (and expectancies) covary and, accordingly, to which behavior can be related. In this sense, Expectancies and Reinforcement Values refer to the different classes of stimuli, or different aspects of the stimulus complex, with which Habits covary.

One further rapprochement of Rotter to Hull is important. Reinforcement Value in many respects corresponds with Hull's Incentive variable (K). The Incentive variable has received relatively little attention in recent Drive theory research; but, at the time when it was added to Hull's system, it was denoted specifically as having to do with motivational (as opposed to learning) phenomena in performance. As elaborated by Rotter, the Incentive value (Reinforcement Value) of a given object or event depends on the situation (stimulus complex) and on the individual's reinforcement history in similar situations. These aspects of Incentive have not been of notable interest to the Hullian group, although such a formulation would not appear to be antagonistic to Hull's model. (Note also that the Incentives of primary interest for Rotter are interpersonal events, rather than physical reinforcing objects.)

In brief, Rotter's model provides a means of discriminating measurable individual differences in higher-order habits which may be related to task performance. Measurement is directed toward individual differences in (1) the value (effectiveness) of the constant incentives in the task situation and (2) the expectancy (subjective probability) that these incentives will occur as a function of incentive-directed behavior.

Further translation of Rotter's system into Hullian terms would not be appropriate here. However, two other variables in Rotter's system should be discussed. Minimal Goal (MG) represents the lowest quantity (or frequency) of a reinforcing event (Incentive) which will be reinforcing for the individual subject. This variable corresponds loosely to "level of aspiration." For example, a grade of A may represent a positive Incentive, but a grade of C or D may be of no Incentive value. The Minimal Goal variable has, for example, been found to be a powerful predictor of academic achievement behavior (ref. 26). Locus of Control (LC) is another Social Learning Theory variable of potential value in this research. Locus of Control is, in effect, a higher-order Expectancy variable which mediates changes, or changeability, of more specific Expectancies. Operationally it corresponds closely to the distinction between contingent and non-contingent reinforcement. The Internal Control pole of the dimension denotes the higher-order Expectancy (experimentally manipulated or otherwise) of 1.00 that the Incentives (reinforcements) of value to the subject are contingent on his own behavior; the External Control pole denotes Expectancy of .00 that his behavior is relevant to incentive attainment. Behavioral change as a function of reinforcement is therefore a function of the subject's higher-order expectancy that reinforcement is contingent upon his behavior. Various individual differences tests of the LC variable have been devised, and results with these tests have been generally favorable.

Rotter's system, in summary, discriminates various parameters of motives (RV, E, MG, and LC) which are confounded or neglected in other approaches to goal-directed behavior phenomena. McClelland, for example, confounds RV and MG phenomena, both conceptually and operationally. In Hullian theory, higher-order habit (H) and Incentive (K) variables can be employed to account for individual differences in goal-directed behavior, but these variables have not yet been articulated and refined for this purpose. Rotter's system provides a first-approximation step toward this goal. The absence of a Drive, or Arousal, variable in Rotter's system, however, makes it essential to integrate Social Learning Theory with Drive (and/or Arousal) theory so that both parameters of motivation can be dealt with systematically and consistently.

Rotter does not specify the needs or motives which should be particularly relevant to performance of any specific task or to behavior in general. He provides no list of motives as did Murray. The rationale for this omission is that such a list can be meaningful only in a descriptive, taxonomic sense, and that no single taxonomic scheme is likely to be optimal for application to all molar behavioral phenomena. The important needs are expected to vary with the subject population, the environmental context, and the specific task (or behavior) in question. The identification of the specific relevant needs or reinforcements must remain an empirical (and inductive) problem for each population, environmental context, task, and behavior.

RELATION OF MOTIVATION TO PERFORMANCE

Simple or Complex?

In common-sense terms it might be assumed that performance improves as a simple direct function of increased motivation. However, this is contradicted by almost all systematic approaches to motivation, and by abundant evidence. The relation of motivation to performance is generally quite complex. Different motivational variables can be expected to have different effects, and the effects of each depend on the specific task employed, on what aspect of performance is considered, on environmental conditions in which performance takes place, and also on other motivational variables. The relation of motivational variables to performance can be expected to be neither additive, nor linear, nor independent of other variables. The complexity of these relationships necessitates a far more elaborate procedure than would otherwise be necessary to construct motivational tests which might predict task performance.

In Drive theory, for example, the relation of Drive to performance is assumed to depend on the dominance of the correct response. Arousal theory suggests a curvilinear relation of Arousal to performance, with optimum Arousal level depending on task complexity. McClelland (and particularly Atkinson) propose that nAch is related to task performance (achievement) only when there are achievement cues relevant to task performance. Rotter proposes that needs and expectancies are related in a multiplicative, rather than an additive, function and that neither will be related to performance if available incentives are below the subject's Minimal Goal for such incentives.

The relationship of any motivation variable to performance can be expected, then, to depend on many other variables. Prediction of performance from a test of that motivational variable must be specific to other relevant aspects of the task and task-setting. In searching for items to measure these motivational variables, these other variables had to be controlled. If the resultant tests were to be of general utility, it was necessary to vary other parameters of the task and the task situation, to determine which tests (or items) would predict performance on which kinds of tasks and in which environmental conditions. It is unlikely that any test (or item) of motivation will ever be capable of predicting performance on tasks in general, across task situations in general.

Variables to Control

From previous literature the following seemed to be the most important categories of variables to control and/or to manipulate in the criterion task setting.

Task Variables

(1) Dominance of Correct Response: From Drive theory it is expected that a measure of individual differences in Drive level should be directly related to performance on any task in which the correct response is dominant, but inversely related to performance whenever some one or more incorrect responses are dominant. Where there is no clear difference in dominance of correct and incorrect responses, there should be no relationship of Drive to performance if performance is defined as frequency of correct response. If performance is defined as amplitude or latency of response, response dominance should be irrelevant, since Drive should increase amplitude and decrease latency of all responses. Also, all other characteristics of task, situation, and subject should be irrelevant to the relationship of Drive to performance, at least in simple classical conditioning tasks.

(2) Task Complexity: Arousal theory is ambiguous as to the precise meaning of "complexity," and the data of the Arousal theorists fail to clarify the question. The definition of task complexity as related to performance remains, then, an empirical problem. It is clear, however, that dominance of correct response is not all that is implied. Since Arousal and Drive appear to be nearly interchangeable concepts, it would appear that any measure of individual differences in Drive would be assumed to measure individual differences in chronic Arousal level. Since both Drive theory and Arousal theory are well supported by systematic evidence (albeit different kinds of evidence), the definition of task complexity and the distinction between task complexity and response dominance are critical. It seemed tenable to propose three separate dimensions of task complexity, any one (or more) of which might mediate the relationship of individual differences in Arousal (Drive) to performance:

(a) Stimulus Complexity: the complexity of the stimulus display in the task, and the necessity of complex perceptual discriminations in task performance

(b) Response Complexity: the necessity of elaborate coordinated effector responses, including mediation of feed-back from effectors, in executing the response

(c) Mediational Complexity: the necessity of complex cognitive mediating processes (r_g - s_g responses in Hullian terms) intervening between stimulus input and response output, as opposed to quasi-automatic, reflexive responding as in a classical conditioning paradigm.

Any or all of these aspects of task complexity may be critical in mediating the relationship of Arousal (Drive?) to performance on various tasks.

(3) Specificity of Sensory Modalities and Effector Mechanisms: Beyond these task parameters it remains possible that some motivational

phenomena may be fairly specific to sensory modality or effector mechanism. For example, increased Drive (or Arousal) might enhance performance on a task involving auditory input but inhibit performance on the same task with visual input, because of differential effects of Arousal on adaptation, habituation, or orienting responses in visual and auditory modalities. Individual differences in acuities or skill in perceptual discrimination in specific modalities may represent more than a random source of error. Barrett (ref. 27), for example, found evidence suggesting that MAS scores may be inversely related to absolute visual thresholds. The same motivational variables may have different effects on performance of a task requiring vigorous large-muscle activity than on a task requiring controlled inhibition of effector response. These illustrations are entirely hypothetical. Such phenomena, however, are quite possible, and could serve to complicate further any attempt to reach generalized conclusions regarding the relationship of some motivational variable to performance.

Variables in the Task Situation

(1) Drive and Arousal Variables: Drive, Arousal, and all other motivational variables are assumed to depend on environmental conditions. If in a given task these variables are related in a curvilinear function to performance, then the effect of individual differences will depend on environmental conditions. Arousal theory would predict, for example, that individual differences in Arousal will be inversely related to performance, in environmental conditions producing greater than optimal Arousal; in less-than-optimal Arousal conditions, individual differences in Arousal should be directly related to performance. Whatever environmental conditions mediate Arousal (strength of noxious UCS, "anxiety" as mediated by instructions or other stimuli, etc) would be expected, then, to interact with individual differences in chronic Arousal level in determining performance.

This particular interaction is not expected in Drive theory, except as amended by the Broen and Storms (ref. 20) hypothesis. However,in Drive theory one might expect similar phenomena as a function of S_D - the Drive - inducing stimulus itself. The addition of S_D to the total stimulus complex alters the stimulus complex, of course; consequently, it may alter the subject's habit hierarchy in that situation and, thus, the relative dominance of alternative responses, including the correct vs incorrect responses. In any case, the possibility of a curvilinear relationship of Drive to performance, and consequently of an interaction between individual differences in Drive level and Drive-inducing conditions of the task setting, cannot be dismissed.

(2) Incentive Variables: Another class of environmental variables to be controlled is the entire range of incentive conditions in the setting. In a task involving a simple, involuntary, reflexive response, incentive factors

may be irrelevant; but in any other task they can be expected to contribute substantially to variance in performance, and to interact with individual differences.

Task-relevant incentives may be of two varieties. First, in a task employing noxious stimulation (eg, electric shock, or an air puff to the cornea) one can assume incentive to escape, avoid, or attenuate the noxious stimulation if possible. Second, the entire class of learned incentives may be relevant: the incentive to do well on a task perceived as representing an important ability (nAch), the incentive to please the experimenter, or at least to avoid his disapproval (need for approval), etc. A complication arises when the incentives may appear to the subject to be associated with performance of the criterion task in some conditions, but associated with performance of a "distractor task" in other conditions. If performance on the distractor task inhibits performance on the criterion task, incentive magnitude may be inversely related to performance of the criterion task.

In reference to all incentives, particularly in reference to social incentives, it can be assumed that there are individual differences in incentive value. Attaining the experimenter's approval is likely to be more important for some subjects than for others, and a constant level of shock may be more noxious for some than for others. Here, again, the environmental incentive conditions may interact with individual differences in incentive value. A low-intensity shock may be of negligible incentive value for some but of effective incentive value for others. If the incentive function of the shock is related to task performance, these individual differences in incentive value may be related to individual differences in performance. A higher-intensity shock may be of sufficient incentive value for all subjects to alter task performance, and if the incentive is adequate for all subjects, then individual differences in incentive value may not be related to differences in performance.

Other incentive-relevant variables may also be distinguished. In the McClelland-Atkinson model, these would include environmental cues which suggest achievement-relevance of the task or arouse achievement motivation, such as the experimenter's describing the task as a measure of intelligence or mental quickness. In Rotter's system these would include variables which affect the subjects' expectancies as to the probability of incentive attainment as a function of alternative behaviors (eg, working hard on the criterion task, or working hard on the distractor task, to avoid disapproval of the experimenter).

(3) Other Environmental Variables: Many other environmental variables might be related to performance on various tasks, and might also interact with such motivational variables. These would include temperature, humidity, lighting conditions, noise, distracting stimuli, etc. There is no priori empirical or theoretical basis for anticipating how these

variables might interact with motivational variables in affecting performance. Their effect, singly and in interaction with motivational variables, must remain questions for subsequent research.

Subject Variables

Beyond the probable interactions of individual differences variables with task and environmental variables, the separate motivational variables may often interact among themselves. For example, the relationship of Drive or anxiety to performance may depend on the subject's level of achievement motivation. For high nAch subjects, those of higher Drive level may perform better on a given task, whereas among low nAch subjects Drive differences may be reflected in behavior not relevant to task performance. Individual differences in incentive value may be unrelated to individual differences in expectancy of incentive attainment (eg, doing well on the task), but for subjects low in either, performance may be poor.

In summary, it is unlikely that one can assume simply, "the higher the motivation, the better the performance." Motivational variables may often be related in curvilinear fashion to performance. More important, the relationship of any single motivational variable to performance may depend on a multitude of other variables, eg, specific characteristics of the task and of the environmental conditions in which the subject performs. These complexities necessitated careful control of the variables with which motivational variables might interact. More important, they required collecting criterion data in a variety of tasks and environmental conditions, to provide some possibility of devising a test battery that would be useful in a variety of settings.

SECTION III
METHODOLOGICAL PROBLEMS

IDENTIFICATION OF VARIABLES

In most traditional procedures to devise an empirically valid test, the methodology for data collection is fairly simple. One starts with the largest feasible number of tests or items, collects test and performance data from the largest feasible number of subjects, and proceeds with the item analysis. This procedure has been effective in devising tests of ability, aptitude, intelligence, etc, and fairly effective in devising attitude scales.

This simple strategy was doomed to fail, however, in constructing a test (or test battery) of motivational variables. The multidimensionality of motivation was one major obstacle. Whereas such variables as intelligence or attitudes may well include discriminably different parameters, these parameters are ordinarily highly related. Different motivational variables, on the other hand, may be totally unrelated to one another.

The more serious obstacle was the interaction of motivational variables with other variables. Performance of one task (or in one setting) might be related to Drive but not to achievement motivation; the reverse might be true on another task. A valid test of a motivational variable might then predict performance differences on some tasks and not on others. For the test battery to be useful, it was necessary to specify which test in the battery was likely to predict differences in performance on which tasks and in which situations, etc. Consequently, the variable represented by each test, and the other variables which determine its relationship to performance, had to be identified. An undifferentiated list of items assessing motivational variables would be minimally useful in predicting performance differences on any specific task, even if each of the items were sometimes capable of predicting performance (ie, depending on task, situation, etc). The test battery, then, had to come with rules for use, and these rules required identifying the variables assessed by each test.

Face validity is a notoriously unreliable criterion for identifying the variable measured by a specific test or item. At least two other procedures are available to increase the accuracy of identifying variables. In increasing order of power, these are:

1. Correlation of the Item with Other Items or Tests Reflecting the Same Variable. This represents a variation of convergent validation (ref. 28), applied here to items rather than to tests. Virtually all existent tests of motivational variables (eg, the MAS, or the nAch scale in the Edwards Personal Preference Schedule) have sufficient internal consistency that items in the scale can be assumed to possess convergent validity of this sort. (Following the convergent validation paradigm, however, it is possible that many items may fail on grounds of discriminant validity. An item in the MAS, for example, may correlate with total scores on the MAS but may correlate even more highly with scores on a social-desirability response-set scale. Thus it is moot whether the item measures anxiety or response-set, or both.)

2. Correlation of the Item with Performance. Often one can predict how a given individual differences variable should be related to performance. For example, in a clearly-defined achievement task, individual differences in need for achievement (and, therefore, tests and items designed to measure this variable) should be directly related to performance. Where there is theoretical or empirical ground to predict such a relationship, the task performance itself can serve as the criterion for identifying the variable measured by the item.

The latter is the more powerful criterion for establishing the meaning of the item and, therefore, for basing assumptions as to the tasks, situations, etc in which the item is likely to predict performance. It is also the more difficult. At present it is still difficult to state with high confidence how any motivation variable should be related to performance on any given task in any given situation. There is still too much inconsistency in available data (eg, discrepancies between Driv[illegible]y and Arousal theory data) to permit confidence in anticipating thes[illegible]hips in any new task or situation.

This problem suggested a pro[illegible]could serve other ends while serving to clarify the relationships o[illegible]ational variables to performance on the task employed. The procedure is a variant of the convergent operations (ref. 29) or convergent validation (ref. 28) strategy. Holding all other conditions constant, one can experimentally manipulate the motivational variable in question. For example, one can systematically vary intensity of noxious stimulation (shock, or air puff) to produce differences among groups in Drive or Arousal. The differences among groups can then indicate the relationship of that variable to performance on that task, in that setting. This can serve as the direct criterion of the results that should be expected concerning the relationship between performance and individual differences on that variable. It provides the generally more powerful and less ambiguous experimental procedure as an external criterion of the effect of that variable on performance. This criterion can be used to interpret results with the more uncertain individual differences test of the same (purported) variable.

This convergent operations strategy is far from new. It has seldom been employed, however, where both experimental and differential operations were involved. Following Campbell & Fiske's (ref. 28) logic that maximum power in convergent validation comes from employing maximally different methods for assessing the same variable, it would seem particularly powerful to combine experimental with differential methods.

The Iowa group has approximated this strategy in studies of eye-blink conditioning. The same hypothesis concerning conditioning as a function of Drive has been tested by employing the MAS and by manipulation of air-puff intensity as measures of Drive. However, these studies have not ordinarily employed both Drive measures simultaneously; MAS and UCS intensity have been investigated in separate studies.

It seemed most efficient and powerful to employ both measures, or operations, in each study. Subjects could be randomly assigned to treatment conditions designed to manipulate the variable in question. Within each condition they could then be categorized (high, middle, low, for example) on the basis of scores on the test designed to measure this variable. The treatment conditions and the categorization could then represent different factors in the data analysis, permitting a factorial analysis of variance design, (with the statistical power associated with such a design. If both operations demonstrated the same relationship to performance, then (a) there would be two independent sources of evidence for the relationship of that variable to performance, (b) there would be evidence of the validity of the test as a measure of that variable, and (c) consequently there would be empirical evidence of the meaning or validity of the separate items within the test. If the experimental manipulation and the test measure disclosed different relationships to performance, one would have to question the validity of the test (and/or, sometimes, of the experimental operation), and remain cautious in assuming what the test (and individual items therein) measures and was likely to predict.

The particular virtue in this procedure, however, involved two other advantages. First, the factorial design would permit increased statistical power in isolating variance associated with either factor. Second, it would permit exploration of possible interactions of individual differences with environmental variables. As discussed previously, it was critical to isolate these interactions, to determine the range of tasks and conditions in which the test battery would be effective.

Consequently it seemed particularly desirable to manipulate conditions of criterion testing along the same dimensions represented (purportedly) by the individual differences tests.

SELECTION OF CRITERION TASKS

As discussed previously the effect of any given motivational variable on performance is likely to depend on the nature of the task, as well as on other characteristics of the task setting. In selecting items against a performance criterion, then, the selection of the criterion task was critical. On it would depend not only which items were selected but also what other tasks they could be used with. Items which predict performance on one task should predict performance on similar tasks. Therefore, it was essential to identify which parameters of the task determined the relationship of the motivational variables to performance, so that one could anticipate what other tasks might be susceptible to effects of these variables.

Many task parameters may be important, such as dominance of correct response, task complexity, receptor and effector mechanisms involved in performance, etc. There is little a priori ground for determining which parameters are most important, or where the criterion tasks should fall on any of these parameters. Of all possible kinds of tasks, however, the most useful might, it seemed, be a classical conditioning task. If it can be assumed that most tasks involve at least some degree of classical conditioning, a relatively "pure" classical conditioning task seemed likely to provide the best chance of selecting items of general utility in predicting performance on a wide range of tasks. These items might not be useful in tasks (a) which do not involve classical conditioning, (b) in which other task variables obscure classical conditioning phenomena, or (c) which are conducted in task-setting circumstances in which classical conditioning phenomena are obscured. They should, however, be at least somewhat useful in a great number of tasks.

One specific purpose of this research was, then, to devise a test which would predict individual difference in classical conditioning. The selection of a classical conditioning task, however, was far from a simple matter. The simplicity of classical conditioning is more apparent than real. The distinction between classical and other kinds of (specifically, instrumental) conditioning is a continuing subject of disagreement (see, for example, refs. 30, 31, and 32). These discussions are generally conducted at a theoretical level, but their ramifications at an operational level, in the selection of a particular task, are considerable.

The following criteria were adopted for selecting a classical conditioning task. These are listed in increasing order of restrictiveness and in decreasing order of agreement among those who have discussed the problem. For the purposes of this research project, a classical conditioning task was defined as one which:

(1) involves a simple, "reflexive" (ie, unlearned or previously conditioned) response to an unconditioned (or previously conditioned) stimulus, the UCS, which response
(2) does not occur to some other stimulus (the CS) prior to conditioning, but
(3) can be elicited by the CS following some number of CS-UCS pairings;
(4) ideally this response (either UCR or CR) should not require cognitive mediation, ie, it should not be a voluntary or even ideally a conscious response;
(5) even more ideally, the response should not be capable of voluntary emission, so that the reflexive nature of the response can be most confidently assumed; and
(6) the response (CR or UCR) should not serve an instrumental (ie, an incentive-attainment) function.

It is the last criterion which was most critical and most difficult in selecting a "pure" classical conditioning task. Spence (ref. 11) has been explicit that the linear relationship of Drive to performance obtains only in incentive-free tasks, which he assumes classical conditioning tasks to be. In the expanded Hullian model, Drive (D) and Incentive (K) are additive, and jointly combine multiplicatively with habit strength (H) to determine reaction potential (E): $E = H \times (D + K)$. The influence of Drive can be discerned only in the absence of Incentive. Spence postulates that Incentive factors are absent in true classical conditioning, so that in such a task $E = H \times D$.

Thus absence of Incentive becomes the most important criterion in defining true classical conditioning.

It is questionable whether any task has yet been devised which meets all these criteria fully and unquestionably. Furthermore, the more nearly a task seems to approximate these criteria, the more likely it will involve severe problems of instrumentation and feasibility. Perhaps the most frequently employed tasks purporting to represent classical conditioning are (1) eyeblink conditioning, (2) GSR conditioning, and (3) cardiac conditioning. The more specific methodological problems of each of these are as follows.

Eyeblink Conditioning

This is the task most employed in Drive theory studies. There are several major problems in its use, however. The task requires both fairly complex apparatus (not immune to apparatus reliability problems) and rather long experimental sessions to establish stable conditioning. Even with fairly long sessions, individual differences tend to be less stable across trials than would be desired in using this as a criterion task. All these problems add up to an inflation of experimental time required.

More important are the failures of this task to meet the above criteria for a classical conditioning task. Whereas the first four criteria are met, the last two are not: (a) the eyeblink response can easily be voluntarily emitted, and such voluntary responses can obscure interpretations of the data, since (b) the noxious stimulus (air puff to the cornea) clearly does provide an incentive for the eyeblink response.

Some procedures have been devised to detect voluntary responses and thereby control for them. It is, of course, difficult to estimate how successful these techniques are. In any case, the presence of incentive cannot be controlled. It is possible that the confounding of Incentive with Drive properties of the UCS is in part responsible for the consistent finding of superior performance with high UCS intensity, with less consistency in finding superior performance by subjects with high MAS scores. There is one further methodological problem in the use of eyeblink conditioning. Operant blink rate is often used as a criterion of anxiety, or Drive level. Differential rate of conditioning of the eyeblink response could, then, be attributable at least in part to differences in operant rate. This phenomenon is not wholly inconsistent with Drive theory; nevertheless, it could represent further ambiguity in the meaning of data relating anxiety measures to rate of eyeblink conditioning.

GSR Conditioning

The GSR response is not so obviously subject to "voluntary" responding as is the eyeblink, although it probably can be emitted voluntarily, at least indirectly. The GSR response also has no obvious Incentive function, although it remains possible that it, or the central neurological response on which it depends, may mediate some Incentive function. More problematic are the feasibility aspects of GSR conditioning. Stable GSR conditioning, like eyeblink conditioning, may require extended experimental sessions; even then, the stability of individual differences in performance leaves much to be desired; apparatus reliability and calibration are major problems; and the GSR response is notoriously sensitive to random stimuli, including "stimuli" with no observable physical referent. GSR responsivity is also highly subject to variation of temperature and humidity, length of time electrodes remain affixed, etc. Also, some subjects fail to give GSR responses even to distinctly noxious stimuli. GSR conditioning would appear to be superior to eyeblink conditioning in approximating the criteria for a classical conditioning task, but the feasibility problems militate against its use in a project requiring the use of the maximum possible number of subjects, especially where stable individual differences in performance are the prime concern.

Cardiac Conditioning

Cardiac conditioning is even less likely than GSR conditioning to permit confounding by voluntary responses. It is likely, a priori, to be as close as any available task to meeting the most stringent criteria for a classical conditioning task. It is, however, even more subject than other tasks to such feasibility problems as the expense and reliability of apparatus, the length of experimental session required, stability of individual differences in performance, and the necessity of using an extremely noxious UCS to elicit the cardiac response. Also, the stimulus and procedural parameters on which cardiac conditioning depends are still ambiguous. This may ultimately become the task of choice in exploring individual differences in classical conditioning. At present, however, it is not suitable as a task for collecting data with a large number of subjects, as was necessary for this project.

As will be explained in the next section of this report, two classical conditioning tasks were devised specifically for this project. The first was a classical escape conditioning task, modeled after a traditional reaction-time task. This task fell far short of maximal approximation of the most stringent definition of a pure classical conditioning task, but provided many methodological and conceptual advantages while still qualifying by a minimum definition of classical conditioning. The second was a finger-retraction conditioning task which was more easily and reliably instrumented than eyeblink conditioning (or GSR or cardiac conditioning) and at the same time provided better control of Incentive factors than did eyeblink conditioning.

ITEM ANALYSIS PROCEDURES

The conceptual and methodological problems discussed above, and the procedural strategies necessary to deal with them, led to a further problem in item analysis procedure. It is desirable in item analysis to accumulate the largest possible number of subjects. This procedure, however, requires that all subjects be tested under identical conditions on the criterion task, or at least that any variation between subjects in task or task-setting conditions be unlikely to affect performance. Any condition which is likely to affect performance and which varies between subjects would produce a major artifact in the item analysis, by yielding items which are merely a function of chance differences between subjects assigned to different conditions. Similarly, the validity of items can easily be obscured if there is substantial variance in performance related to test conditions.

Since it was necessary, however, to vary task and task-setting variables systematically, the traditional item analysis procedures were not feasible. These variations in task and task-setting variables restricted the number of subjects tested under identical circumstances.

This problem, however, was the least severe of those confronting the research. The "sequential analysis" item analysis procedure (ref. 33) was specifically designed to be used with small numbers of subjects (even of the order of 10-15), and has been shown to be nearly as powerful as the more traditional procedures employing much larger groups of subjects. This procedure is somewhat time-consuming, since it does not lend itself to computer,analysis but its result is little if any less powerful and stable than more traditional procedures.

SELECTION OF TESTS TO EXPLORE

To construct a test battery, it is ideal to explore the greatest possible number of tests and test items. The number of available tests of motivational variables is legion. However, because of restrictions in the amount of time-per-subject available, it was necessary to choose carefully among available tests.

In making this choice, tests were discarded which require extensive time to administer or score (eg, McClelland's TAT technique and French's Insight Test), assuming that these would be of least use to subsequent investigators. As the second step, it was decided to focus on those tests which (a) had the most encouraging evidence of probably validity and (b) measured variables which were most likely to be related to performance on simple, controlled laboratory tasks.

The first test selected was the Taylor Manifest Anxiety Scale (MAS) as a measure of Drive. Although there are many variations of the MAS, and other tests purporting to measure Drive differences, none has shown as much evidence of validity as the MAS. The next two tests selected were the Marlowe-Crowne Social-Desirability (ref. 34) Scale (M-C SD), as a measure of "need for approval," and Liverant's Locus of Control (LC) (ref. 35) Scale. Each of these has shown particular power in predicting performance in laboratory tasks and each of these grows out of Rotter's Social Learning Theory, which renders results with these tests more easily interpretable. Need for approval seemed likely to predict effort expended on laboratory tasks in which other incentives are minimal. It has been found to be related, for example, to individual differences in verbal conditioning and perceptual defense (ref. 34). Locus of Control, as defined in Social Learning Theory, should serve as a mediator of the effects of incentives on performance. To the extent that incentive effects were important in any task, LC was expected to be related to performance.

The Edwards Personal Preference Schedule (EPPS) was selected next. This test purports to measure 15 different needs, following Murray's model. Although there is only limited evidence of validity of these scales, the test did provide a pool of 225 additional items for item analysis. Next selected was an adapted Adjective Check List (ACL), simply because it included over 300 items and could be administered in 10 to 15 minutes. In the absence of other tests with prior evidence of ability to predict laboratory task performance, these two tests seemed to be as likely candidates as any.

At various stages of data collection, other tests were employed as time and circumstances permitted. In one study, the Heineman Forced-Choice Anxiety Scale was substituted for the MAS to determine whether the partial control of response-set factors by the Heineman test might provide a more powerful test of Drive. The data analysis did not suggest any such superiority. The Heineman test involves a complex format which does not lend itself to feasible item-analysis procedures. Consequently no item analysis was performed, nor was the test used in other studies. Similarly, the Nowlis Mood-Adjective Check List, used in one study, failed to indicate validity (or predictiveness), and (since it also is refractory to item analysis) was discarded.

Three other tests were employed in some of the studies. One was Rotter's Incomplete Sentences Blank (ISB). This, of course, is not an objective test with discrete items amenable to item analysis. It was employed only as a post-hoc hypothesis-seeking (or variable-seeking) device. The intention was to isolate high and low performers on each task and then scan the ISB protocols to see whether the latter would suggest any variables (or specific responses) which might discriminate the two groups, and which might be capable of being assessed by a tailor-made objective scale.

The criterion data, however, were too complex to permit this procedure, so this effort was abandoned. The other two tests -- Costello's Ego-Involvement Scale (EIS) (ref. 36) and Barratt's Impulsivity Scale (IS) (ref. 37) - were added in the final four data-collection projects, in part simply because there was sufficient available group-testing time to allow collecting more test data. Each of these was recently-devised and is of as yet indeterminate validity, but with some suggestion of promise. Also, each purports to measure a variable which might be expected to relate to performance differences.

SECTION IV
MOTIVATIONAL CORRELATES OF REACTION TIME

RATIONALE AND METHOD FOR REACTION-TIME STUDIES

Rationale

The first task developed for the research was an escape-conditioning task in the model of traditional reaction-time tasks. It only minimally met the a priori criteria for a "pure" conditioning task. No UCR was involved; the response required some conscious mediation, at least initially. Incentive conditions affected performance greatly. However, the task was one which is often used in laboratory research, especially in investigations of the effect of stress on performance, and did include some features of classical conditioning.

The specific rationale for use of this task was that it permitted independent manipulation of (1) experimentally-induced Drive, (2) dominance of correct response, and complexity of (3) stimulus input, (4) mediation processes, and (5) overt response. Consequently it seemed ideal for investigating the extent to which motivational variables interact with these task parameters in affecting performance.

Basic Method

Apparatus

The same apparatus was employed in all five reaction-time studies (see ref. 38 for details and schematic drawings). The response panel was a metal box, 12 by 10 by 8 cm, on the table at which the subject sat. Two neon lights, 8 cm apart on the top of the box, were the stimulus lights.

(These could be replaced, if desired, by apparatus which could present more complex stimulus displays.) The two response keys were toggle switches, placed directly beneath the stimulus lights on the top of the response box. The subject's thumbs were placed against the front of the panel, with his right forefinger on the right response key and his left forefinger on the left key. A pinching movement of thumb and forefinger closed the response key. A red foreperiod cuelight was placed on the table above the response panel.

Three separate tasks could be programmed with this apparatus. In the simple task, only the stimulus light on the right side of the panel was used, and pressing the right response key (under the light) extinguished the stimulus light and the UCS. In the disjunctive task, the two stimulus lights were presented in random sequence on successive trials, and the correct response was to press the key under whichever light was illuminated. In the complex or cross-disjunctive task, stimuli again varied randomly but the correct response was to depress the key not under the illuminated light. These three tasks represented decreasing dominance of the correct response.

To manipulate Drive, an induction coil, energized by four 1.5-volt dry-cell batteries, was employed. The buzz of the coil when energized was of approximately 40 db intensity. For shock conditions (high Drive), the shock was conducted from the coil to the subject by two polished brass electrodes, secured to the subject's leg by an elastic belt. The electrodes were positioned to deliver a tetanizing shock to the gastrocnemius muscles. For no-shock conditions (low Drive), the electrodes were not attached to the subject, so that the only UCS was the buzz of the induction coil. Timers were set to provide a 2-millisecond interval between onset of CS (light) and onset of UCS (shock and buzz, or buzz only). In pilot testing this delay was found to be great enough to assure that the subject would respond to the CS rather than to the UCS, and short enough that the UCS could still not be completely avoided.

Response latency was recorded to the nearest 0.01 sec. Timer, CS, and UCS were all extinguished concurrently by the correct response. If the subject pressed the wrong key, both CS and UCS (and the timer) continued until this key was returned to its original position and the correct key depressed, and this was recorded as an error.

Four timers controlled duration of foreperiod (1, 2, or 3 seconds, in random order), presentation of CS, intertrial interval (5 to 15 sec, randomly varied), and presentation of UCS.

Procedure

Before beginning, the experimenter explained the apparatus and told the subject that the purpose of the study had to do with reaction time. Then the subject attached the electrodes to the calf of his left leg. He was given a series of shocks, beginning at 2.5 volts and increasing in 2-volt steps to 8.5 volts, to assure him that the shock would be tolerable and not dangerous. He was then asked if he were willing to continue. All subjects agreed to continue, although all agreed the shock was extremely unpleasant. Subjects in the no-shock condition did not undergo this procedure.

The instructions were then read (see ref. 38). The subject was told that the UCS would follow the CS by a few seconds but could not be avoided completely, and that both CS and UCS would terminate when the correct response had been made.

Three demonstration trials were given for each task to be used in the study in which the subject was participating. Then testing began.

STUDY I

Purpose

The first study was primarily a pilot study. Aside from refining methodology, the purpose was to further explore the effect of Drive on performance at different levels of dominance of the correct response. In this study there was no exploration of individual differences in Drive, but only of Drive as mediated by a noxious UCS.

Method

Apparatus

The apparatus was as described as above.

Subjects

A total of 20 male undergraduate students, volunteers from a men's residence hall or from an introductory psychology course, served as subjects. Age range was 18 to 25. All subjects reported being right-handed and of normal vision. No subject was told of the use of electric shock before he entered the experimental room.

Experimental Design

Subjects were randomly assigned to one of four treatment conditions, with five subjects in each condition. Treatment conditions represented shock and no-shock and two orders of testing. All subjects were tested on all three tasks (simple, disjunctive, and complex), half in order of increasing complexity and half in reverse order. These two orders comprised the second treatment factor. There were 50 trials on each task, grouped for statistical analysis into 5 blocks of 10 trials each. Trial blocks were treated as a repeated-measures factor in the analysis of variance. The three levels of task complexity (dominance of correct response) comprised another repeated-measures factor. The experimental design, then, was a 2x2x5x3 factorial design (shock, order, blocks, task), with repeated measures over the last two factors. Each subject was tested under one shock condition, in one order condition, across five trial blocks on each of the three tasks.

Procedure

The basic procedure described above was followed. Subjects were given three demonstration trials on each of the three tasks before testing began. On the disjunctive and complex tasks, there was an equal number of right- and left-hand responses, in random order. There was a 30-sec rest period between trial blocks. To maximize response competition, there was no rest period between the last trials on one task and the first trials on the next. This transition occurred without warning, and the subject had no knowledge of the order in which he was to take the three tasks.

Results

Frequency of errors and response latency were the dependent variables. To normalize the latency scores, the total time in milliseconds for each subject on each trial block was transformed to a speed measure by the transformation $X^1 = (1/X)$ times 1000. These scores were subjected to analysis of variance. For the error scores, the extreme heterogeneity of variance within groups and the skewness of the data militated against analysis of variance. The Kruskal-Wallis H test, corrected for tied observations, was used.

The summary of analysis of variance of speed scores is in table I, Figure 1 illustrates results with the speed and error measures.

<u>Speed</u>

Speed decreased significantly as dominance of correct response decreased (tasks main effect, $p<.01$). The differences between the disjunctive task and the complex task supported the assumption that the tendency to press the key under the light was stronger than the tendency to press the key not under the light; the dominant response, then, was correct on the disjunctive task but incorrect on the complex task.

Shocked subjects were faster than nonshocked subjects (shock main effect, $p<.05$). However, this was true only on the simple task. The shock x tasks interaction was significant ($p<.01$); analyses of simple effects confirmed that differences between shocked and nonshocked subjects were significant only on the simple task; contrary to expectation, shock did not inhibit speed on the complex task, except for a slight trend on the first trial block. Also contrary to prediction, shock tended (although not significantly) to inhibit speed on the disjunctive problem, despite the fact that the dominant response was correct on this task.

Speed increased across trials (trials main effect, $p<.01$) in a "learning curve" function, supporting the assumption that conditioning does take place on these tasks. Rate of improvement was greatest on the simple task (task x trials interaction, $p<.05$).

Subjects were faster on the simple task if they had previously had the complex and disjunctive tasks (order 2) than if they had the simple task first (order 1); order did not affect speed on the disjunctive or complex task (task x order interaction, $p<.01$). There appeared, then, to be some warmup or practice effect which enhanced speed on the simple task.

TABLE I

SUMMARY OF ANALYSIS OF VARIANCE OF SPEED[1] SCORES, STUDY I

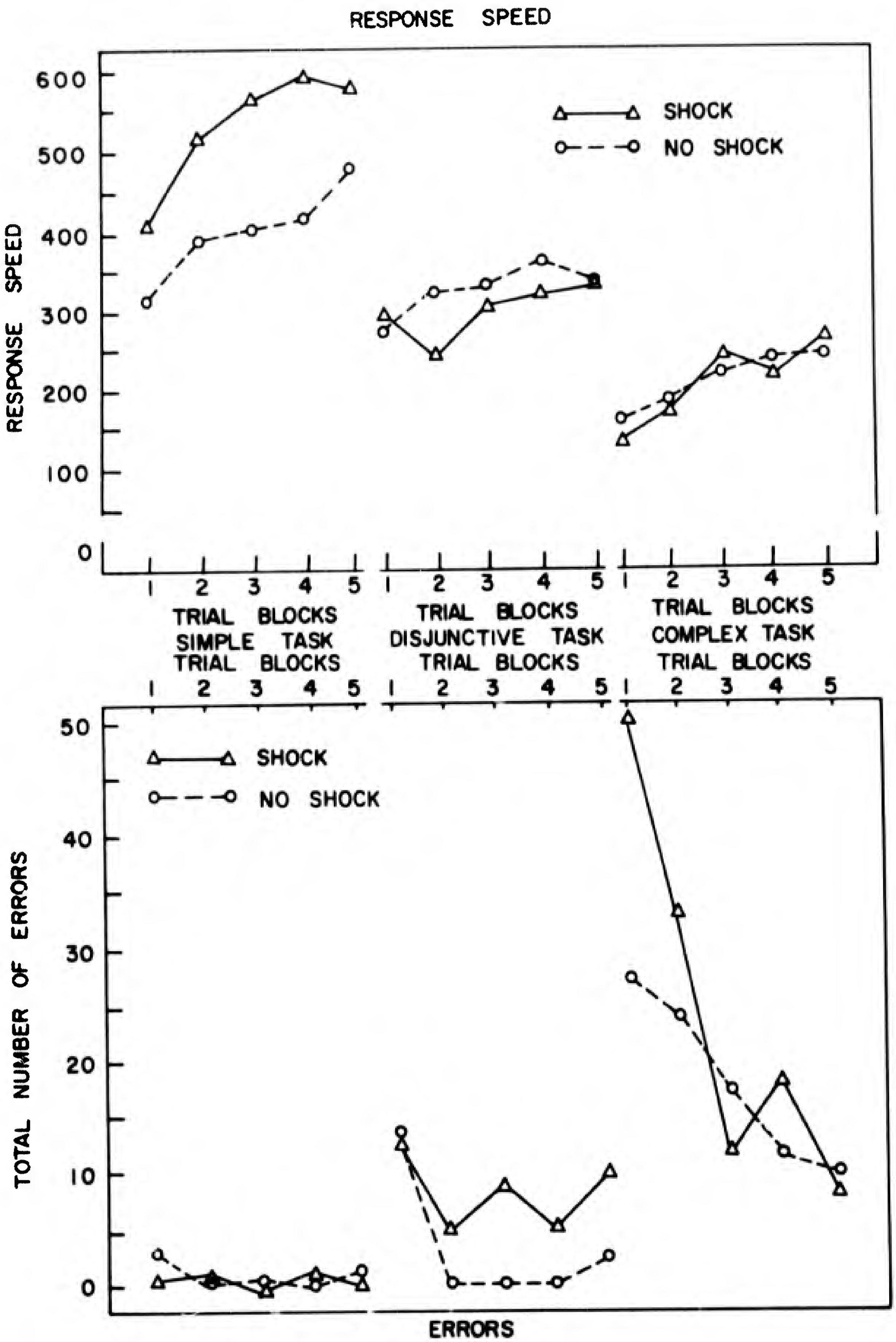
Source of Variation	df	Mean Square	F
Between Subjects	19		
Shock (S)	1	108,499.59	7.74*
Order (O)	1	58,943.17	4.36
S x O	1	4,033.69	1
(Subj. within Grp. Error)	(16)	13,501.91	
Within Subjects	280		
Trial Blocks (Tr)	4	106,256.75	61.45**
S x Tr	4	2,433.16	1.41
O x Tr	4	1,855.84	1.07
S x O x Tr	4	6,122.90	3.54*
(Tr x Subj. within Grp. Error)	(64)	1,729.28	
Tasks (Ta)	2	1,713,683.75	429.65**
S x Ta	2	150,687.21	38.78**
O x Ta	2	75,188.81	18.85**
S x O x Ta	2	3,136.87	1
(Ta x Subj. within Grp. Error)	(32)	3,988.54	
Tr x Ta	8	9,204.65	3.77*
S x Tr x Ta	8	4,987.88	2.04
O x Tr x Ta	8	1,862.54	1
S x O x Tr x Ta	8	4,022.24	1.65
(Ta x Tr x Subj. within Grp. Error)	(128)	2,444.21	

[1]Speed = (1/total reaction time in milliseconds per trial block per subject) x 1000

* $p<.05$
**$p<.01$

Figure 1

Mean Response Speed and Total Error Frequency per Trial Block as a Function of Shock and Task Complexity, Study I.



In order 1 (simple, disjunctive, complex) speed increased regularly from one trial block to the next, on each task, for both shocked and nonshocked subjects. Subjects tested in the reverse order, however, showed a less regular improvement across trials; shocked subjects had a performance decrement on the last trial block, and nonshocked subjects on the third trial block with improvement only on the second and the fifth block of trials (shock x order x trials interaction, $p < .05$). The importance of this phenomenon is minimal.

Errors

Shocked subjects made more errors than nonshocked subjects, on both the complex ($H = 12.29$, $p < .001$) and the disjunctive ($H = 6.39$, $p < .02$) task. There were virtually no errors on the simple task in either condition. There were no significant differences ($p > .50$) in error frequency between order 1 and order 2.

Figure 1 illustrates the relationship of error frequency to shock and trial blocks on each task. On the disjunctive task, shock differences are not apparent on the first trial block. Thereafter, errors virtually extinguish for nonshocked subjects, but are only somewhat reduced for shocked subjects. Shock differences in error frequency, then, occur only after the first trial block on this task. On the complex task, on the other hand, shock differences appear only on the first two trial blocks, and not thereafter (except for apparently random irregularity).

Conclusions

These results generally confirmed expectations. Shock improved performance (speed) on the simple task, but inhibited performance (increased errors) on the complex task. These tasks appeared, then, to be sensitive to the effects of Drive on performance when (1) the correct response is dominant and (2) an incorrect response is dominant. The tasks seemed ideal to use in collecting criterion data for constructing a test of individual differences in Drive level.

There were some discrepancies in the results, however. The most important was the failure of shock to affect speed on the complex task. It had been expected that shock would inhibit speed on this task. Except for a slight trend on the first trial block, it did not. Consequently, even though the results confirmed predictions, the predictions were confirmed with different criteria for the simple and the complex task. This suggested that other variables (the nature of which later became clear) were involved in addition to Drive and response dominance. It also suggested, discouragingly, that different tests might have to be developed for predicting different aspects of performance.

Other discrepancies were less important. Shock did not reduce errors on the simple task; but errors were too infrequent to permit significant effects. Results with the disjunctive task did not suggest a facilitatory effect of shock; the trend, in fact, was opposite. However, the dominance of the correct response in this task may have been too slight to permit Drive facilitation of performance. Oscillatory response potentials and other extraneous variables could have produced the instability of results.

In summary, the results were encouraging, but suggested further investigation of variables other than Drive which affected performance on these tasks.

STUDY II

Purpose

The second study was designed to clarify some of the questions raised by Study I and to begin exploration of individual differences in performance on this type of task.

The questions concerned the complex task. Results with the simple task were clear-cut. The disjunctive task was of lesser interest than the other two tasks, because of its in-between nature. Dominance of correct response was apparently not clear-cut, and the task did not represent any clear point on a continuum of task complexity or of dominance of correct response. The complex task, however, did offer promise of being an optimal task for exploring Drive effects in conditions of dominance of incorrect response. Compared with other tasks (eg, finger mazes, verbal mazes, paired-associates learning) used for this purpose, the complex reaction-time task offered unique advantages. With it, it was possible to hold constant the stimulus and response and still make the correct response either clearly dominant or clearly nondominant. The role of response dominance in mediating Drive effects could, then, be determined with minimal ambiguity. This task also permitted independent manipulation of stimulus complexity, response complexity, and mediational complexity.

Two major questions remained, however. First, why did shock not affect speed on this task? Second, if performance were carried to asymptote, would the correct response become dominant and shocked subjects ultimately excel on both speed and error measures? The latter would be expected both from Drive theory and from the traditional assumption that motivational effects are apparent in asymptotic performance level, rather than in rate of learning. In Study I it was obvious that 50 trials were not enough to approach asymptote or, consequently, to answer the latter question. As for the first question, it was possible that the effect of shock on speed was obscured by other variables such as (1) the small number of subjects, (2) the order factor, or (3) individual differences in Drive among subjects assigned to the same treatment conditions.

There was one further question. In the first two trial Blocks, errors were far more frequent among shocked than among nonshocked subjects, yet the speed scores of the two groups were almost identical. As the task was instrumented, the response was not recorded until the subject had pressed the correct key. If he had previously or simultaneously pressed the wrong key, the timer continued until that key was returned to original position and the correct key pressed. When an error occurred, additional time was required to correct the error and make the correct response. Error trials, then, showed generally poorer speed scores than nonerror trials. Since shocked subjects had more error trials during the first two trial blocks, they may have been responding more quickly (but more often incorrectly) than nonshocked subjects. Whether there was a significant difference between groups in speed of first response (whether or not correct) during the first two trial blocks could not be answered from Study I results, since the only speed scores recorded were for occurrence of correct response.

The second study was designed to answer these questions, and particularly to control the sources of contamination which might have obscured Drive differences early in performance.

Method

Apparatus

The apparatus was not altered, except (a) only the complex task was programmed and (b) a second timer was added and wired to record latency of first response.

Subjects

The total number of subjects was 30. Subjects were volunteers from introductory psychology classes, each of whom had been given the MAS and also the M-C S.D and LC.

Experimental Design

All subjects were tested for 150 trials (15 blocks of 10 trials each) on the complex task. Half were tested under shock conditions and half were tested under noshock conditions. Within each treatment group, subjects were trichotomized on the basis of MAS scores (high, middle, low). This resulted in a 2x3x15 factorial design (shock x MAS x trial blocks), with trials as a repeated-measures factor. Dependent variables were frequency of errors, latency of correct response, and latency of first response.

Procedure

There was no change in procedure except for alterations permitted by omitting the simple and disjunctive tasks.

Results

Table II includes summaries of analyses of variance of speed of correct response and of error frequency. There was a high correlation between speed of first response and speed of correct response (see table III). Equivalent results were found with these two measures. Discussion will be limited, then, to speed of correct response, except to note that this measure was apparently not seriously contaminated by the additional time required to correct errors when errors occurred.

Speed

Shock failed to have a significant overall effect, as in Study I. However, there was a significant ($p < .01$) shock x trials interaction (see figure 2). Shock increased speed on trial blocks 2 through 7. Although shock slightly reduced speed on the first trial block (as in Study I), it consistently increased speed thereafter until performance approached asymptote. At asymptote there was no clear effect. Speed was inversely related, however, to anxiety ($p < .01$); and this was particularly true during early trial blocks (anxiety x trials interaction, $p < .01$). Again, however, this relationship was not clear on the first trial block, when middle-anxious subjects were fastest (see figure 2).

TABLE II

SUMMARY OF ANALYSES OF VARIANCE OF SPEED AND ERROR SCORES, STUDY II

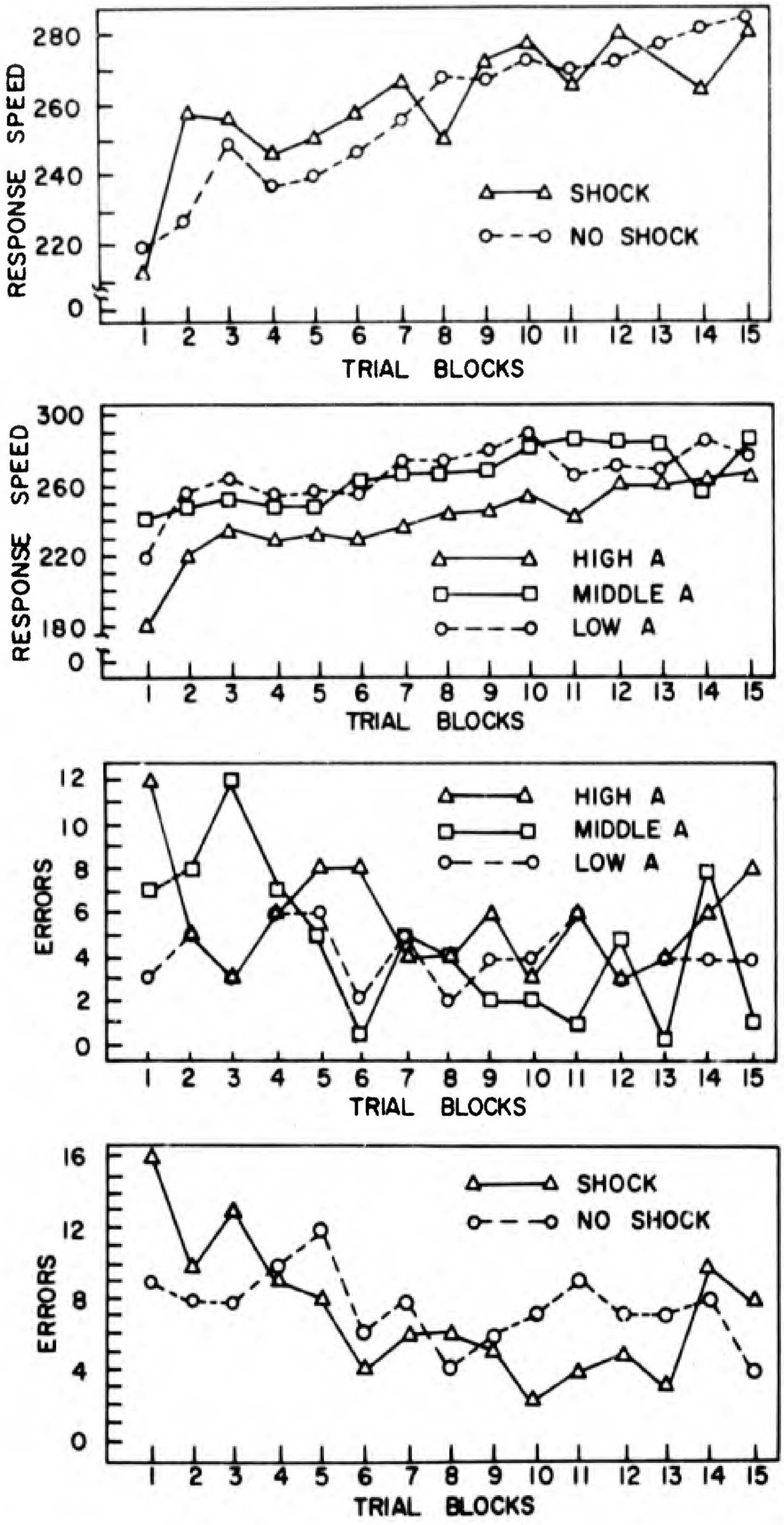
Source of Variation	df	Speed Mean Square	Speed F	Errors Mean Square	Errors F
Between Subjects	29				
Shock (S)	1	0.07788	1	0.00000	1
Anxiety (A)	2	1.92889	9.5**	1.48667	2.5
S x A	2	1.00969	5.0*	1.64667	2.8
(Subj. within grp. Error)	(24)	0.20266	—	0.58556	—
Within Subjects	420				
Trial Blocks (Tr)	14	1.72756	24.6**	0.91333	2.3**
S x Tr	14	0.60988	8.7**	0.78571	2.0**
A x Tr	28	0.33355	4.7**	0.42714	1.0
S x A x Tr	28	0.37256	5.3**	0.41095	1.0
(Tr x Subj. within grp. Error)	(336)	0.07035	—	0.39508	—

* $p < .05$
** $p < .01$

The middle-anxious subjects were also deviant in their response to shock. Shock reduced speed of both high- and low-anxious subjects, but increased speed of middle-anxious subjects (anxiety x shock interaction, $p < .05$), especially during early trial blocks (shock x anxiety x trials interaction, $p < .01$). In the no-shock condition anxiety was inversely related to speed, but in the shock condition middle-anxious subjects were slightly faster than low-anxious subjects, with high-anxious subjects slowest of all. The implication of this interaction is unclear, except to suggest that (a) the relationship of anxiety to performance on this task is not always linear or (b) more likely, the MAS is not a pure, linear measure of anxiety, but contaminated by other variables. The latter possibility is more apparent in results discussed below.

Figure 2

Mean Response Speed and Total Errors per Trial Block as a Function of Shock and Anxiety, Study II.



(A DENOTES ANXIETY)

Errors

Error frequency decreased across trials ($p. < .01$), and decreased more rapidly in the shock than the no-shock condition (shock x trials interaction, $p < .05$). As in Study I, shock increased errors during early trial blocks, but not thereafter (see figure 2).

There was a trend ($p < .25$) for error frequency to be directly related to anxiety. This was especially clear on the first trial block although the anxiety x trials interaction was nonsignificant (see figure 2). Shock tended to increase errors of middle-anxious subjects, but decrease errors of high- and low-anxious subjects (shock x anxiety interaction, $p < .10$). High-anxious subjects made more errors than low-anxious subjects in both the shock and the no-shock condition; in the no-shock condition errors were least frequent for middle-anxious subjects, however, and in the shock condition errors were most frequent for middle-anxious subjects.

Correlational Data

Speed and error measures were correlated with M-C SD and LC scores as a subsidiary interest, since these test scores were available. These correlations produced the most important results (see table III). Error frequency was inversely related ($r = -.51$) to M- C SD, or need for approval, test scores in the shock condition. M-C SD scores were directly related ($r = .44$) to speed in the no-shock condition. MAS scores were directly related to errors in both conditions. There were also trends ($p < .10$) for LC scores to be inversely related to both speed and errors in the shock condition, and MAS scores to be inversely related to speed in both conditions.

Conclusions

Two questions were answered clearly. The speed of correct response measure was not seriously contaminated by time required to correct errors; and at asymptotic performance levels shock (and anxiety) have little effect.

Much more important results were obtained than these, however. Shock increased errors and increased speed on early trial blocks. The latter phenomenon was not apparent in Study I. Anxiety, like shock, was directly related to errors. But anxiety and shock had opposite effects on speed. Whereas shock increased speed on early trial blocks, anxiety reduced it.

TABLE III

INTERCORRELATIONS OF INDEPENDENT AND DEPENDENT VARIABLES, STUDY II

Correlations with Speed of Correct Response

Measure	Groups Shock	No-Shock	All Subjects
Speed, First Response	.96***	.83***	.87***
Errors	-.05	-.06	-.06
MAS	-.32*	-.26*	-.27*
M-C SD	-.08	.44**	.25*
LC	-.26*	.04	-.11

Correlations with Number of Errors

Measure	Groups Shock	No-Shock	All Subjects
MAS	.43**	.37*	.37*
M-C SD	-.51***	-.08	-.23*
LC	-.29*	.01	-.08

* p<.10
** p<.05
*** p<.01

The effect of Anxiety on speed was consistent with Drive theory expectation. The effect of Shock was not.

The correlational data suggested an explanation for this reversal. In the no-shock condition, need for approval enhanced speed; in the shock condition need for approval had no effect on speed, but reduced errors. Need for approval can be assumed to be an incentive-type, rather than a Drive-type, motivational variable. In retrospect it was obvious that shock also served an incentive (for speed), as well as a Drive, function. In the absence of a shock incentive, the experimenter's approval is presumably the primary incentive for maximum speed. With a shock incentive, there is obviously adequate incentive to respond as quickly as possible, regardless of the subject's evaluation of the importance of the experimenter's approval. In the shock condition, however, need for approval still was an incentive to be careful to avoid errors.

The shock incentive for speed could, then, have counteracted the Drive effects of shock. Ironically, the MAS measure of Drive may have been more valid than the experimental manipulation. With the MAS measure, results were perfectly consistent with Drive theory expectations.

These results also suggested another methodological problem inherent in the task. Since need for approval reduced errors in the shock condition, it appeared that need for approval increased the subject's cautiousness. This caution was presumably reflected in a greater tendency to inhibit responses, for some period of vicarious trial and error, to avoid error responses. The responses available to the subject must have included, then, more alternatives than simply pressing the left key or pressing the right key; they must have included also the response of inhibiting, if briefly, both lever-press responses. The effects of Drive on errors, and to some extent on speed, should depend on the relative strength of the inhibiting response; pressing the key under the light might not always be the dominant response tendency. The shock incentive must, in fact, alter the relative strengths of inhibiting and overt responses. The manipulation of dominance of correct and incorrect (overt) responses was not as clear, then, as had been assumed.

Despite these problems, the results confirmed Drive theory expectations more clearly than those of Study I. Drive as measured by the MAS had the expected effects on performance. The unexpected increase in speed on early trials associated with shock could be ascribed to the selective incentive function of shock. The task appeared, then, to be sensitive to individual differences in Drive as expected, and also sensitive to individual differences in an incentive-type variable, need for approval. It seemed only somewhat sensitive to the effects of differences in locus of control.

STUDY III

Purpose

The primary purpose of this study was to control response-inhibition, if possible. One function this could serve was to clarify the effects of response-inhibition on performance. Another was to reduce the degree to which other variables were contaminated or obscured by this variable. Instructions were changed to reduce individual differences in response-inhibition. Half the subjects were told that the most important aspect of performance was speed; the other half were told that minimizing errors was most important.

A second purpose was to investigate a different test measure of anxiety. The MAS had been found correlated with the M-C SD in Study II; and the contamination of the MAS with social desirability factors has been widely suspected. Contamination of anxiety measures with individual differences in social desirability response set would be doubly unfortunate in this research. First, it would obscure the effects of anxiety and Drive. Second, it would defeat the purpose of distinguishing the nature of the different motivational variables independently related to performance. The Heineman anxiety scale was substituted, then, for the MAS. This is a multiple-forced-choice anxiety scale, rather than a true-false test like the MAS; it was designed as an attempt to control response sets. A second purpose of this study, therefore, was to evaluate the Heineman test as an anxiety, or Drive, measure.

The third purpose was to return to investigation of the simple task. Individual differences in Drive were not investigated in Study I. It was intended to determine whether the effects of individual differences in Drive are parallel to those of shock and to replicate the earlier findings regarding shock.

Method

Apparatus

The same apparatus was employed as before, this time programmed to present the simple and complex tasks. The apparatus for recording latency of first response was removed.

Subjects

Volunteers from introductory psychology courses were used; total number of subjects was 120. All subjects had been tested previously on the M-C SD, LC, and Heineman Anxiety Scale.

Experimental Design

Half the subjects were tested with shock and half with no-shock. Half were given instructions emphasizing the importance of speed and half were instructed to be as careful as possible to avoid errors. All subjects performed both the simple and the complex tasks, with 50 trials on each task. Half the subjects had the simple task first, half the complex. Within each of the eight resultant treatment conditions the 15 subjects were trichotomized on the basis of Heineman Anxiety Scale scores, to represent an individual differences factor in the analysis of variance. The experimental design represented a 2x2x2x3x5 (shock x instructions x order x anxiety x trial blocks). Results on the simple and complex tasks were analyzed separately. Dependent variables were number of errors (for the complex task) and latency of correct response (for both simple and complex tasks).

Procedure

There was no alteration in the procedure of Study I and Study II except for the amendment in instructions to emphasize the importance of response speed or error-avoidance and amendments necessitated by employing both the simple and the complex tasks. In the instructions all subjects were told that both speed and error-avoidance were important - but that one or the other was "especially important."

Results

Complex Task: Latency

It was not necessary to transform latency scores. Figures and analyses represent latency of correct response, then, rather than speed. The summary of analysis of variance of latency scores (summed across the 10 trials of each trial block for each subject) is in table IV.

Instructions had no clear overall effect. They did, however, mediate the effects of other variables in several interactions, as discussed below.

As in Study II, shock reduced latency ($p<.01$), but latency was directly related to anxiety on the first trial block (anxiety x trials interaction, $p<.01$). High-anxious subjects were consistently slower than low-anxious subjects, but middle-anxious subjects were again inconsistent. They were intermediate between high- and low-anxious subjects only on the first and last trial blocks; on the second trial block they had the lowest latency scores of all subjects, and on the third and fourth trial blocks the highest latency scores. Still, the trend ($p<.25$) was for latency to be directly related to anxiety overall, as in Study II.

TABLE IV
SUMMARY OF ANALYSES OF VARIANCE, STUDY III

Source of Variation	df	Complex Task Errors F	Complex Task Latency F	Simple Task Latency F
Between Subjects	119			
Shock (S)	1	10.25**	9.82**	24.45**
Anxiety (A)	2	1	1.88	1.98
Instruction (I)	1	1	1.03	1.74
Order (O)	1	1	1.53	9.44*
S x A	2	1.39	1	1
S x I	1	1	1	1
S x O	1	1	1	1
A x I	2	2.08	1	2.43
A x O	2	1	1	1.35
I x O	1	1	1	1
S x A x I	2	1	1	1.35
S x A x O	2	1.63	1	1
S x I x O	1	1	1	1
A x I x O	2	2.22	1.92	5.14*
S x A x I x O	2	1	1.94	1.48
(Subj. within grp. Error)	(96)			
Within Subjects	480			
Trial Blocks (Tr)	4	73.70**	62.16**	120.15**
S x Tr	4	1	1	1.05
A x Tr	8	1.55	3.08**	2.86**
I x Tr	4	1.34	1	1
O x Tr	4	2.32	4.28**	30.75**
S x A x Tr	8	1	1	2.13*
S x O x Tr	4	1.90	2.08	4.60**
S x I x Tr	4	1	1	1
A x I x Tr	8	1.91	1.89	4.09**
A x O x Tr	8	1	1	1
I x O x Tr	4	1	3.13**	1
S x A x I x Tr	8	1	1	1
S x A x O x Tr	8	1.22	1	1
S x I x O x Tr	4	1	1	1
A x I x O x Tr	8	1.24	1.93	1.57
S x A x I x O x Tr	8	1.12	1.43	2.47*
(Tr x Subj. within grp. Error)	(384)			

* $p < .05$
** $p < .01$

Instructions had little effect in order 1, in which the complex task was preceded by the simple task. In order 2, speed instructions increased latency on the first two trial blocks and decreased latency only on the final trial block (instructions x order x trials, $p<.01$). Apparently (1) the effects of instructions were obliterated by the time the subject had completed the simple task and (2) overemphasis on speed may actually inhibit speed on early trials.

Performing the simple task first (order 1) reduced latency on the complex task for the first three trial blocks, but had the opposite effect on later trials (trials x order interaction, $p<.01$). In order 1 performance reached asymptote at the third trial block; in order 2 no asymptote had been reached by the end of all 50 trials. This again suggested, as in Study I, a practice or warmup effect which transfers from one task to another, and also an eventual fatigue or boredom effect.

In summary, shock again reduced latency, as in Study II, and anxiety again increased latency on early trials, although not as clearly and consistently as in Study II. Middle-anxious subjects again reacted atypically, this time to the instructions rather than to shock. The instructions employed did not clearly affect latency, and in fact tended to have effects opposite to those expected (on early trials of order 2, in which instructions would be expected to have their clearest effects).

Complex Task: Errors

The summary of analysis of variance of errors on the complex [illegible] also in table IV.

As in studies I and II, shock increased errors ($p<.01$) and err[illegible] decreased across trials ($p<.01$). No other variables had significant effects.

Three interactions, however, nearly reached .05 significance. These were anxiety x instructions x trials, order x trials, and shock x order x trials. Of these the first was most important. Error frequency was directly related to anxiety on the first trial block, as in Study II. However, this trend was slight (total errors on first block were 116, 114, and 107 for high-, middle-, and low-anxious subjects, respectively); it did not obtain on later trial blocks; and it was affected by instructions. Middle-anxious subjects were again deviant. They made more errors on the first trial block under accuracy instructions, and fewer under speed instructions, than did high- or low-anxious subjects. After the first trial block anxiety differences were inconsistent.

As for the other two interactions, error frequency decreased more slowly in order 2 than in order 1, again suggesting a warmup effect, although order did not affect errors on the first trial block. Also, shock failed to increase errors on the first trial block when the complex task was performed first; on all later trial blocks, and on all trial blocks in order 1, shock increased error frequency.

In summary, shock again, as a Drive stimulus and incentive for speed, increased errors. The corresponding effect of anxiety was minimal. Accuracy instructions failed to have the expected effect of reducing errors. And middle-anxious subjects again reacted atypically, by making relatively more errors under accuracy instructions and fewer under speed instructions than other subjects.

Simple Task: Latency

Table IV also includes the summary of analysis of variance of latency scores on the simple task. There were several significant interactions, including a significant ($p<.05$) interaction of all five variables (anxiety x shock x instructions x order x trials). Many of these interactions were both exceedingly complex and of limited or obscure conceptual significance. Only the clearly important results will be discussed and illustrated. A table of means (table V) is included so that the other interactions can be examined if desired.

Shock reduced latency ($p<.01$), as in Study I. However, there was no parallel effect of anxiety, except for a slight trend on the first trial block; after the first trial block the middle-anxious subjects were fastest, and high-anxious subjects slowest (anxiety x trials, $p<.01$). The overall trend ($p<.25$) was for latency to be greatest for high-anxious and least for middle-anxious subjects. In the shock condition, this ordering of anxiety groups was consistent across all trials; in the no-shock condition speed on the first trial block was directly related to anxiety (parallel to the general effect of shock), after which middle-anxious subjects were fastest (anxiety x shock x trials, $p<.05$; see figure 3). Instructions also affected the differential rate of improvement of different anxiety groups (anxiety x instructions x trials, $p<.01$). This effect, also illustrated in figure 3, was quite complex.

Instructions had different effects on different anxiety groups when the simple task was performed first (anxiety x instructions x order, $p<.05$): speed instructions <u>reduced</u> speed of middle-anxious subjects, although other subjects reacted as expected. When the complex task had been performed previously, all anxiety groups had slightly higher latencies under speed than under accuracy instructions. Apparently the effects of instructions were

TABLE V

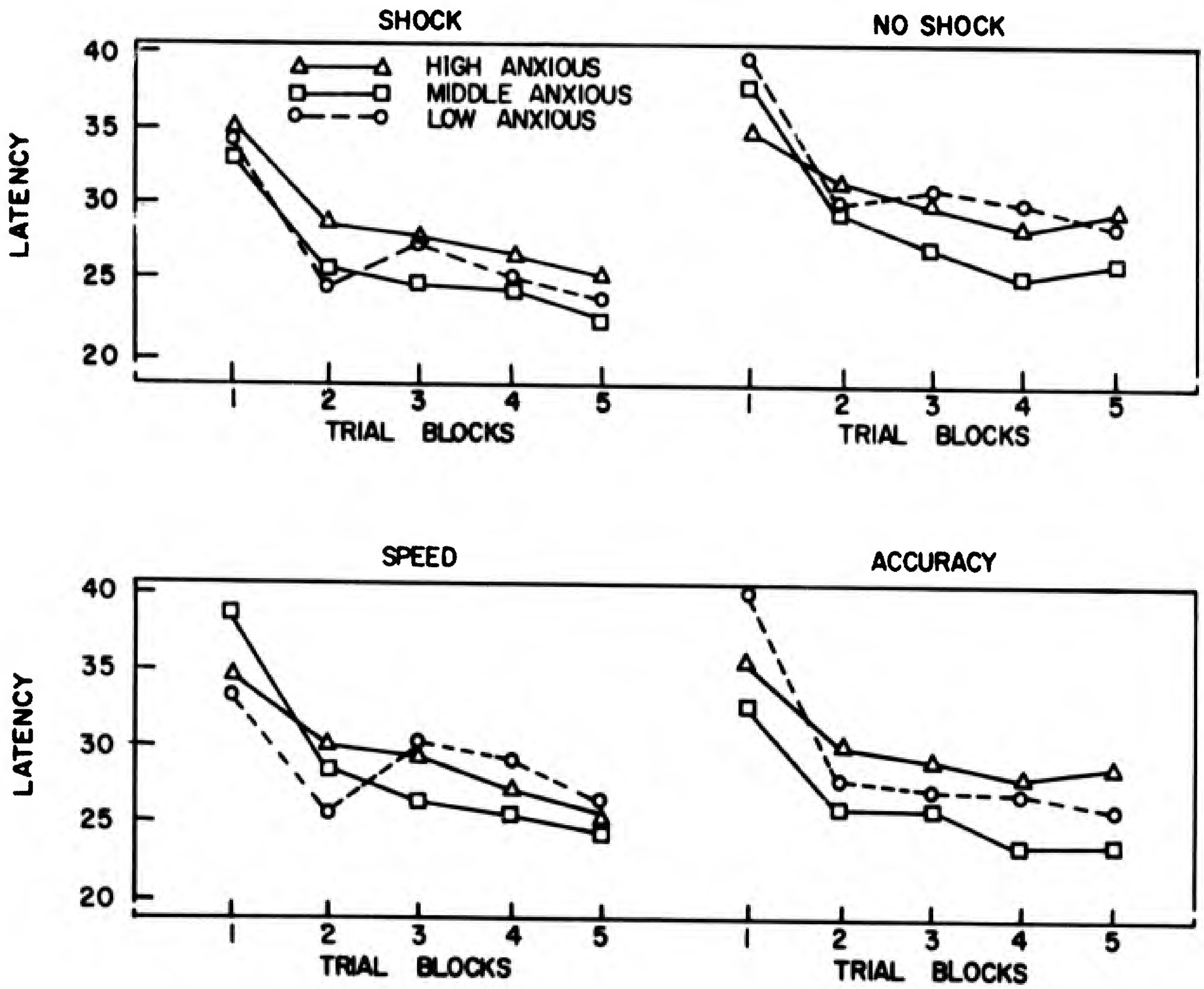
MEANS[1], RESPONSE LATENCY, SIMPLE TASK, STUDY III

			Speed Instructions			Accuracy Instructions		
			Anxiety Groups			Anxiety Groups		
			High	Middle	Low	High	Middle	Low
ORDER 1	Shock	1	140	155	123	151	125	187
		2	133	123	111	138	117	145
		3	129	119	117	140	127	130
		4	134	142	132	125	119	128
		5	127	118	109	121	119	132
	No Shock	1	148	204	182	163	167	204
		2	149	155	137	156	122	160
		3	146	155	166	167	125	146
		4	133	129	144	144	120	151
		5	140	141	142	168	121	146
ORDER 2	Shock	1	220	218	158	187	171	215
		2	148	138	125	145	121	112
		3	153	129	166	130	119	127
		4	141	115	131	128	106	110
		5	121	112	128	132	98	102
	No Shock	1	185	191	202	204	188	201
		2	169	138	144	160	157	140
		3	146	122	150	146	147	146
		4	127	127	163	151	129	145
		5	135	123	142	146	140	138

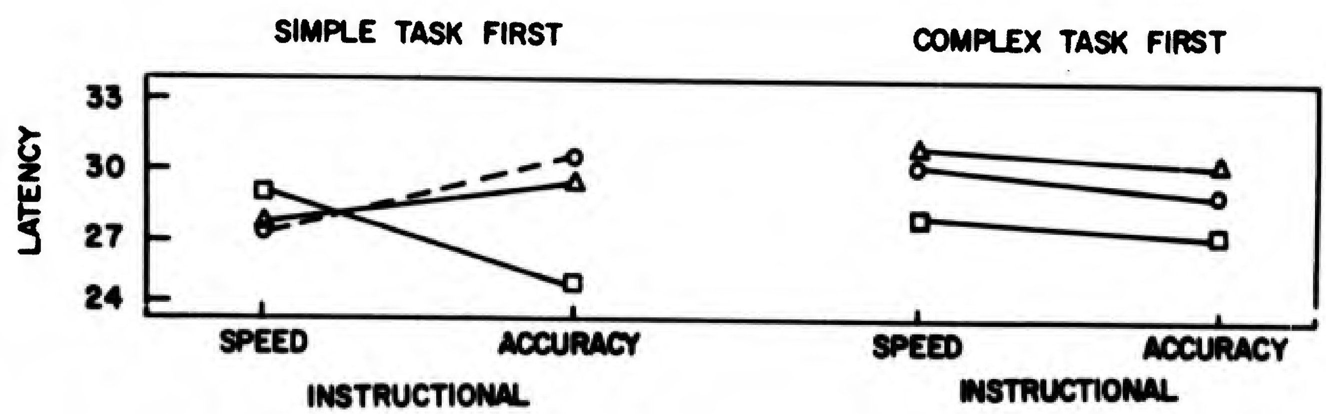
1 cell means = mean latency in milliseconds summed across 5 subjects per cell

Figure 3
Mean Latency of Different Anxiety Groups, Simple Task, Study III

Latency per Trial Block in Different Conditions



Latency as a Function of Task Order and Instructions



obliterated by the time the complex task had been completed. Again, it may be noted that the middle-anxious subjects were deviant in their response to instructions.

Although latencies generally decreased across trial blocks ($p < .01$), the rate of improvement varied with order of testing (trials x order, $p < .01$), and shock had differential effects on rate of improvement in different testing orders (shock x trials x order, $p < .01$). These phenomena were of limited conceptual importance.

Simple Task: Errors

As in Study I, there were too few errors on the simple task to permit statistical analysis. Of the 120 subjects 40 had no errors, and 33 erred only on the first trial of the simple task when it was preceded by the complex task (order 2). Since the subject was given no warning of the transition from one task to the other, these errors simply reflect failure to anticipate the transition. A total of 128 errors were made by the 120 subjects. Beyond the errors which reflected failure to anticipate transition to the simple task in order 2, the majority of errors occurred on the first trial following a rest period; this suggests that some subjects tended to anticipate (mistakenly) when the transition was going to occur.

Conclusions

Instructions had no clear, general effect in manipulating dominance of response inhibition vs uninhibited responding. The several significant interactions including instructions, however, indicated that the instructions were not irrelevant. Different subjects reacted differently to the instructions; specifically, middle-anxious subjects reacted little, and sometimes atypically. Effects of instructions were different on the task performed first than on the task performed second. Instructions, then, mediated effects of other variables. It is also possible that they partially succeeded in controlling response dominance; this is suggested by the obvious statistical power apparent in these data analyses, manifest in the numerous interactions of minimal order which were statistically significant.

It should be noted that the instructions were not as pointed as they might have been to manipulate response dominance. Differential emphasis was placed, alternatively, on the importance of maximum speed or of avoiding errors; but the subject was not instructed to disregard either altogether. Stronger emphasis might have had clearer effects.

The effects of shock observed in studies I and II were replicated. Shock reduced latency on both the simple and the complex task, and increased error frequency on the complex task.

The effects of anxiety in the complex task observed in Study II were also generally replicated. Anxiety again increased latency, especially on early trials, and again tended (very slightly) to increase error frequency on early trials. Anxiety and shock had opposite effects on latency in the complex task, and it was the anxiety, rather than the shock, effect which was consistent with Drive theory expectation. The incentive function of shock, however, might still have outweighed to Drive function in affecting latency.

On the simple task, however, anxiety tended again to increase latency; and this is contrary to Drive theory expectation. This suggested that "anxiety," as measured by anxiety tests, may reflect a general dominance of response inhibition, as much as or more than it reflects a heightened Drive state. This would not be wholly inconsistent with clinical impressions regarding anxiety.

Other results clearly suggested that scores on the anxiety test are contaminated by non-Drive variables. Specifically, middle-anxious subjects were often deviant, rather than intermediate between low- and high-anxious subjects; and this often occurred where no reasonable rationale would suggest a curvilinear relationship of Drive to performance. These results were often suggested of a "deviant response set." Paradoxically, however, the deviance was manifest by the modal third, rather than by either extreme third, of the population.

Whatever other ramifications these were, it was clear that the Heineman anxiety scale showed little evidence of greater power or validity than the MAS, or of lesser contamination by non-Drive variables. The Heineman scale was dropped from the research therewith, and the MAS retained as the measure of individual differences in Drive.

STUDY IV

Study IV was performed solely to collect criterion data for test development. Subjects were tested on both simple and complex tasks, 50 trials each. Half were tested under shock and half under no-shock conditions; in each group half were tested on the simple task first and half on the complex task first. The 72 subjects, volunteers from freshman AFROTC classes, were assigned randomly to these four conditions. They had previously taken, in their AFROTC classes, the MAS, M-C SD, LC, EPPS, ACL, IS, and EIS. Latency scores were computed for each subject on the simple and complex tasks separately, and errors on the complex task were tabulated. Those three performance measures were correlated with MAS, M-C, SD, LC, IS, and EIS scores; they were also correlated with scores on each of these tests, plus the EPPS and ACL, based on only those items selected in item analyses, with items keyed as suggested by the item analyses. These correlations are reported in the final section of this report, along with correlations from data obtained in other studies.

The data collected in this study were examined for any unique results; but their were none. In general, the previously determined effects of shock, anxiety, and task were replicated.

STUDY V

Purpose

The purpose of this final reaction-time study was to investigate the effects of an incentive other than shock. The combined, and possibly counteracting, Drive and incentive effects of shock had still not been entirely clarified. The primary interest in Study V was to compare shock with some experimentally manipulated incentive which would be minimally Drive-producing. In this study individual differences were not of direct concern.

Method

Apparatus

The same apparatus was used as in previous studies, with the addition of a signal light, with timer, as explained below.

Subjects

A total of 40 volunteers from freshman AFROTC classes served as subjects.

Experimental Design

Subjects were randomly assigned to one of four conditions: control, shock, light control, and light incentive. These conditions are explained below. There were 10 subjects randomly assigned to each condition. There were 4 trial blocks (10 trials per block) each on the simple and complex tasks; all subjects were tested on both tasks. Order was counterbalanced but not treated as a factor in the analysis of variance. A separate 4 x 4 (conditions x trial blocks) analysis of variance was performed for the simple and the complex task. The dependent variable was mean latency of correct response per trial block; transformation of latency scores was not necessary.

Procedure

Subjects had 20 practice trials on each task. On the basis of these practice trials, the subject's mean latency was computed for the simple and the complex task, as explained below. During the subsequent experimental trials, this mean latency was used to determine the interval between onset of the stimulus and onset of incentive. No incentive or shock was employed for subjects in the control condition. In the shock condition the incentive was shock delivered in the same manner and intensity as in previous studies except for the adapted interval between stimulus and shock. In the light-control condition, a red light was illuminated if the subject failed to respond within the predetermined interval; subjects were told this was "just an added factor" of no importance. In the light-incentive condition subjects were told that the light represented the average reaction time of Air Force personnel, as determined from previous research. It was assumed that this would, then, represent an achievement incentive for speed.

The interval between stimulus and light (or shock) for each subject was determined as follows. All practice trials on which errors occurred were excluded, as were all trials on which latency was greater than 500 milliseconds, on the complex task, or 350 milliseconds, on the simple task. The mean latency was computed for remaining trials. The interval was set at 20 milliseconds below this mean for the complex task and 10 milliseconds below the mean for the simple task.

Results

Preliminary analyses of variance confirmed that there were no significant differences among groups on the practice trials, on either the simple or the complex task.

Table VI summarizes analyses of variance on the experimental trials. Latencies on error trials were included. A separate analysis, excluding error trials, produced equivalent results.

Latencies differed significantly ($p < .01$) among incentive conditions on the simple task. On the complex task these differences were nonsignificant. Means for the four conditions, on each task, are in table VI. On each task latencies were least in the light-incentive condition. Newman-Keuls tests for differences among means revealed that the difference between the control and the light-incentive condition was significant ($p < .05$) and between the light-control and light-incentive condition nearly significant ($p < .10$) on the simple task.

Conclusions

The light incentive reduced latency on both tasks, although the effect was nonsignificant on the complex task. This renders ambiguous the findings in previous studies regarding the effect of shock on latency. Clearly the shock was an incentive for speed as well as a Drive stimulus. Whether the shock effect on latency in the simple task is a combination of incentive and Drive, or attributable only to incentive, cannot be determined. On the other hand, the unexpected effect of shock in reducing latency on the complex task can, in retrospect, be viewed as wholly an artifact of incentive. Drive effects were obviously obscured by incentive effects.

It is also important that the light incentive reduced latency more than shock did, on both tasks. Even though these differences were nonsignificant, it appears that the shock may have had some disrupting effect on performance. It is scarcely likely that the shock was of lesser magnitude as an incentive than was the achievement-related light.

Finally, it is notable that latencies tended to be less in the light-control than in the control (no light) condition. It is possible that the subjects defined the light as an incentive, or standard to meet, for themselves, even though this was not suggested to them. This conclusion must be guarded, because of the slight and nonsignificant differences. However, it is supported by the fact that incentive conditions had more effect on the simple than on the complex task. The complex task is no doubt more difficult, and therefore more challenging, than the simple task. It may, then, have more built-in achievement incentive; and other incentives, then, may have less effect.

TABLE VI

SUMMARY OF ANALYSES OF VARIANCE AND MEANS OF LATENCY SCORES, STUDY V

Analysis of Variance

Source of Variation	df	Complex Task MS	Complex Task F	Simple Task MS	Simple Task F
Between Subjects	39				
Incentive Conditions (I)	3	168.575	1.35	407	4.57**
(Subj. within grp. Error)	(36)	(124.938)	-	(89.03)	-
Within Subjects	120				
Trial Blocks (Tr)	3	38.092	1.95	203.8	40.92**
I x Tr	9	17.908	1	8.68	1.74
(Tr x Subj. within grp. Error)	(108)	(19.528)	-	(4.98)	-

** $p<.01$

Means

Condition	Complex Task	Simple Task
Control	44.55	32.33
Shock	42.68	27.98
Light-Control	42.93	29.53
Light-Incentive	39.60	24.68

SUMMARY: MOTIVATION AND REACTION TIME

Motivational Variables in Reaction Time

At least four classes of variables can be identified which jointly determine performance on reaction-time tasks. These are (1) incentive variables, (2) Drive or Arousal, (3) distractor or disrupting variables, and (4) task variables.

Incentive Variables

Of the motivational variables, incentive variables seem to have the most profound effects. Shock, as it was used in these studies, clearly was an incentive for speed. The tasks were in effect escape-conditioning tasks; the reinforcer for the lever-press response was escape from shock. Shock may well have had Drive-inducing, and also disrupting, functions; but these functions could not be separated, in these studies, from the incentive function. Only if shock were delivered independent of response speed could these variables be disentangled. In future research concerning Drive effects, this procedure would be preferable to that used in these studies. Incentive effects may contaminate results of investigations of Drive effects on any task when Drive is induced by a noxious stimulus which is also an incentive for some particular response.

At least two social incentive variables were identified. These were achievement incentives and the approval of the experimenter. The latter, in Study II, seemed partly to counteract the effect of the shock incentive. The achievement incentive in Study V seemed at least as potent as the shock incentive. Social incentives may, then, be very important determinants of performance on tasks like these. This represents a major challenge in future research. It may be extremely difficult to control all social incentives which the subject might attribute to the task situation. On the other hand, these social incentives would be very important objects of study in themselves. Other social incentives, in addition to achievement and approval, are likely to be of importance. Their isolation, measurement, and control remain to be accomplished.

Separate incentives seem to have additive effects, at least up to the physiological limit of the subject to perform. Shock and achievement incentives had less effect on the complex than on the simple task; this may simply imply that the complex task is more challenging, and has some degree of built-in achievement incentive. However, both tasks must offer at least some achievement incentive, to most subjects; and still, additional incentives do improve performance.

In future research, the additivity and asymptotic aspects of incentives should be carefully explored. It is apparent that incentives vary on more than a present-absent continuum. It is also apparent that, in addition to Minimal Goal phenomena, there are effective-maximum phenomena.

Drive and Arousal

Little can be concluded, from these studies, about the effects of Drive or Arousal on reaction-time performance. Shock manipulation of Drive was contaminated by incentive effects; and anxiety tests, as measures of Drive level, must be interpreted with caution. Anxiety was generally related to increased latency on both the simple and the complex task. On the simple task, this is difficult to rationalize with theoretical expectation. Middle-anxious subjects were often deviant, rather than intermediate between high- and low-anxious subjects. This, too, suggests that scores on the anxiety tests are contaminated by non-Drive variables. High-anxious subjects, as defined by test scores, may tend to have a generalized response-inhibition tendency. Middle-ranking anxiety test scores may be contaminated by some peculiar response-set characteristic, analogous to a deviant response set, which is related to task performance in some way. Low-anxiety test scores are quite likely to be contaminated by social desirability response-set.

One implication of these problems is that, when anxiety tests are used as Drive measures, it is unsafe to compare only high- and low-scoring groups, as has traditionally been done. The inclusion of the middle third in these studies was accidental; it was done only because of the ultimate purpose of item analysis and test construction. It now seems quite likely that anxiety tests are not linear measures of anxiety or Drive. In addition, Drive may not always have a linear relation to performance. Inclusion of a middle-scoring group is necessary to avoid inappropriate conclusions drawn from groups defined by anxiety test scores.

One further point should be made about Drive or Arousal effects. It is possible that, in most laboratory studies with humans, the range in Drive levels is actually negligible. Most subjects are generally alert, attentive, etc. Additional Drive stimuli may increase this state only slightly, in terms of the total imaginable range of Drive or Arousal levels. These variables may indeed have the postulated relationships to behavior; but in the laboratory, they may vary across too narrow a range to have major effect. Similarly, anxiety tests may in fact be valid measures of Drive, but the difference in Drive level between the most and the least anxious subject may be relatively little. In future research a major effort should be made to investigate Drive effects by experimental manipulation, but by methods uncontaminated by incentive or other variables. Anxiety-test scores will remain a useful supplement to experimental manipulation. Drive effects will remain in doubt, however, until there is some convergence between the experimental and differential measures which cannot be easily attributed to non-Drive variables. A bootstraps strategy seems to be necessary.

Distractor and Disrupting Variables

In Study V it appeared that shock may leave some disrupting effect, as well as Drive and incentive effects, on performance. This was implied by the finding that the light-incentive condition produced shorter latencies on the simple task than did shock. This difference was not statistically significant. Still, it suggests care that S_D effects of Drive-inducing stimuli not be overlooked. Any detectable stimulus, particularly a noxious stimulus, elicits some response, virtually by definition. This response may not be objectively detectable. Nevertheless it may serve to interfere with, or at times to enhance, another response which is the object of study. Experimental manipulation of Drive is to that extent further complicated.

Task Variables

These studies concerned only one task variable: the relative strengths of correct and incorrect response tendencies, as they mediate effects of motivational variables. Even this variable may have been contaminated by the differential challenge, or inherent achievement incentive, in the two tasks. It was also clear that response-inhibition competed with both overt response tendencies, so that specification of the dominant response was compromised. Neither of these problems necessarily vitiates the advantage of these tasks in exploring relative response strengths. Both, however, suggest further task variables to investigate. Ultimately a very broad range of task variables must be explored before generalized conclusions can be reached regarding the effects of motivational variables on performance.

Implications for Test Development

In light of the findings of these studies, it would be optimistic to expect tests to predict reaction-time performance very accurately. These studies contributed substantially toward isolating variables which should be controlled and manipulated while collecting criterion data. Unfortunately, these variables became apparent only as a result of the research. Criterion data collected in these studies were, then, subject to limitations which were only being discovered at the time.

Approval incentives and achievement incentives do, however, seem importantly related to performance, at least in some conditions. Individual differences in the importance of these incentives should be measurable. A start toward this objective was made. Without question, further work will be required to develop very powerful measures of these variables and to define the conditions in which they are related to performance. This work should be profitable.

The data from these studies will be of less value in developing a measure of individual differences in Drive level, simply because Drive effects are more uncertain. Drive differences may also be less related, even in a pure case, to reaction-time performance.

SECTION V
MOTIVATIONAL CORRELATES OF FINGER-RETRACTION CONDITIONING

RATIONALE AND METHOD FOR FINGER-RETRACTION TASK

Rationale

In the reaction-time tasks, classical conditioning phenomena were obviously obscured by incentive effects and higher-order habit variables. Also, Drive effects was minimal, or at least uncertain. These problems made the tasks of little use for exploring Drive phenomena, for developing tests of individual differences in Drive, or for developing tests which would be likely to predict performance in classical conditioning tasks. A task was sought which would better represent a "pure" classical conditioning task.

A cardiac conditioning task was tried in pilot studies; but feasibility problems, for the purposes of this research, seemed insurmountable. Ultimately a finger-retraction conditioning task was devised which seemed to meet immediate needs, and also offered some unique advantages over other classical conditioning tasks. Specifically, it was possible to investigate Drive and incentive effects (of shock) separately, without altering other aspects of the task. Since incentive effects were so prominent in the reaction-time studies, this offered a chance to clarify what role they had in obscuring Drive phenomena.

The rationale for this task, then, was (1) that it more nearly represented a clear classical conditioning task and (2) would permit separation of Drive and incentive phenomena. Specifically the task did involve a reflexive UCR, which did not require conscious mediation or voluntary emission; it was possible, through CS-UCS pairing, to elicit this response with a CS; and neither UCR nor CR served any obvious incentive function. In Arousal theory terms the task represented a clear minimum of task complexity. It did not, however, permit manipulation of the dominance of correct and incorrect responses, as did the reaction-time tasks.

Method

Apparatus

The apparatus included an audio oscillator, ear phones, two induction coils, shock electrodes, a microswitch, timing mechanisms, and recording devices (see ref. 15 for details). The subject was seated in a one-arm classroom chair in a normally illuminated room. Stimulus controls and recording equipment were located in an adjoining room.

The subject's right arm and wrist were strapped to the arm of the chair. The UCR and CR (finger retraction) were recorded by a sensitive microswitch clamped to the arm of the chair, adjusted so that normal pressure of the forefinger in an extended resting position held it down (closed).

The CS was a buzzing sound of approximately 61 db SPL intensity generated by an induction coil wired in parallel with a 6 volt power source. (Ambient room noise with equipment running averaged 35 db SPL.) The UCS (electric shock) was generated by an identically wired induction coil. For subjects tested under avoidance conditions (incentive present) one silver electrode (1 cm diameter) was placed on the tip of the microswitch and the other (3 by 1 3/4 cm silver GSR electrode) was taped to the palmar surface of the forefinger. Withdrawal of the middle finger within the CS-UCS interval permitted the subject to avoid the shock. For the classical condition (incentive absent) both electrodes were taped to the fingers; the small silver electrode was taped to the middle finger and the larger GSR electrode was taped to the forefinger. This arrangement did not permit the subject to escape the shock even though he responded (retracted the middle finger) within the CS-UCS interval.

A timer was started with the onset of the CS and was stopped when the microswitch was released by movement of the finger. Latency of response was read directly to the nearest .01 seconds.

The CS was presented 0.4 seconds before the UCS and terminated with the end of the UCS, which lasted 0.2 seconds. During test trials the shock circuit was broken by a knife switch wired in series with the subject and the shock source. CS-UCS interval, time between trials, and duration of the UCS were controlled by means of five timers.

A 40 db SPL, 200 cps tone, generated by an audio oscillator, was delivered by earphone to the subject's left ear to mask the sound of recording equipment.

Procedure

To disguise the nature of the task and to prevent voluntary responding, subjects were told that the purpose of the experiment was to test their ability to concentrate under stress. They were asked to count the number of times that they heard a buzzing sound (the CS) and were told that to make the task difficult they would receive shocks from time to time (the UCS). They were told not to voluntarily remove their fingers from the switch, but were cautioned not to resist any natural movement of the finger. All subjects seemed to believe the instructions and responded accordingly. Actually, the counting task was rather difficult; at the end of the experiment very few subjects were able to give an accurate count.

Following instructions, preliminary tests were conducted to test for pseudo-conditioning. Five shocks, of the intensity used during conditioning, were presented alone and followed by five presentations of the CS alone. There was no evidence of pseudo-conditioning: none of the subjects responded to the CS.

After these preliminary tests, 5 blocks of 50 trials each were presented, for a total of 250 trials. Within each block of 50 trials, 10 test trials (in which the shock was not presented) were given. The intertrial intervals were randomly varied from 5 to 15 seconds. One-minute rest periods were given between blocks.

STUDY VI

Purpose

Prior to this study pilot studies had been conducted to determine optimal CS and UCS intervals and intensities, etc, and to ascertain that finger-retraction conditioning could be obtained with this general method. The purposes of Study VI were (1) to explore the effects of incentive and of Drive (both UCS-induced and MAS-measured) on performance on this task and (2) to provide data for developing a test of individual differences in Drive level.

Method

Apparatus

The apparatus was exactly as described above.

Subjects

Volunteers were recruited from three sophomore AFROTC sections; total number of subjects was 120. All subjects had been previously tested in their respective classes on the MAS, M-C SD, LC, EPPS, ACL, ISB, and a specially-devised sentence-completion test.

Experimental Design

Subjects were randomly assigned to treatment conditions. Half were assigned to the avoidance (incentive present) condition and half to the classical (incentive absent) condition. In each group, half were assigned to the high shock and half to the low shock condition, to manipulate Drive, and Drive-plus-incentive, magnitude. (In the low shock condition, shock values ranged from 30 to 70 m. amps, never exceeding the latter; in the high shock group, values ranged from 75-125 m. amps, never less than the former. Within each condition current was adjusted for each subject to a level which at least elicited a UCR but was within the subject's pain tolerance; current was readjusted throughout testing to compensate for adaptation, ie, to maintain UCRs at approximately constant vigor.)

Within each of the resultant four treatment conditions subjects were trichotomized on the basis of MAS scores. All subjects were tested through 50 test trials (five blocks of 10 trials each), randomly distributed among 200 reinforced trials in which the UCS was presented. The result was a 2 x 2 x 3 x 5 (conditioning procedure x shock strength x anxiety x trial blocks) factorial analysis of variance design, with repeated measures across blocks.

Procedure

The procedure was identical to that described above.

Results

A response on a test trial was considered a CR if it occurred within 410 milliseconds after onset of the CS. For each subject two conditioning scores were obtained for each block of 10 test trials: (1) frequency of response and (2) mean latency of response on test trials in which one or more CRs occured. For those blocks of trials in which no CRs occurred within the specified time period, a maximum mean latency score of 410 was assigned. Latency scores were transformed to log units to normalize the distribution. Mean log response latency was, as expected, inversely related to frequency of response. Summaries of analyses of variance of latency and frequency measures are in table VII. Results are illustrated in figure 4.

Incentive significantly increased frequency and reduced latency of responses ($p < .01$), as did the higher shock intensity. Both incentive and Drive increased performance, then, as expected; incentive effects, however, were of far greater magnitude. Response frequency increased regularly across trials, following a learning function; the decrease in latency across trials was nonsignificant. In general, the frequency measure appeared more powerful than the latency measure.

Shock effects on frequency were greater for the avoidance than the classical procedure (conditioning x shock, $p < .05$), but there were no such effects regarding latency. Conditioning was more rapid in the avoidance than in the classical procedure (conditioning x trials, $p < .05$ and $p < .01$). Conditioning was also more rapid under high schock than under low shock (shock x trials, $p < .01$ and $p < .01$).

TABLE VII

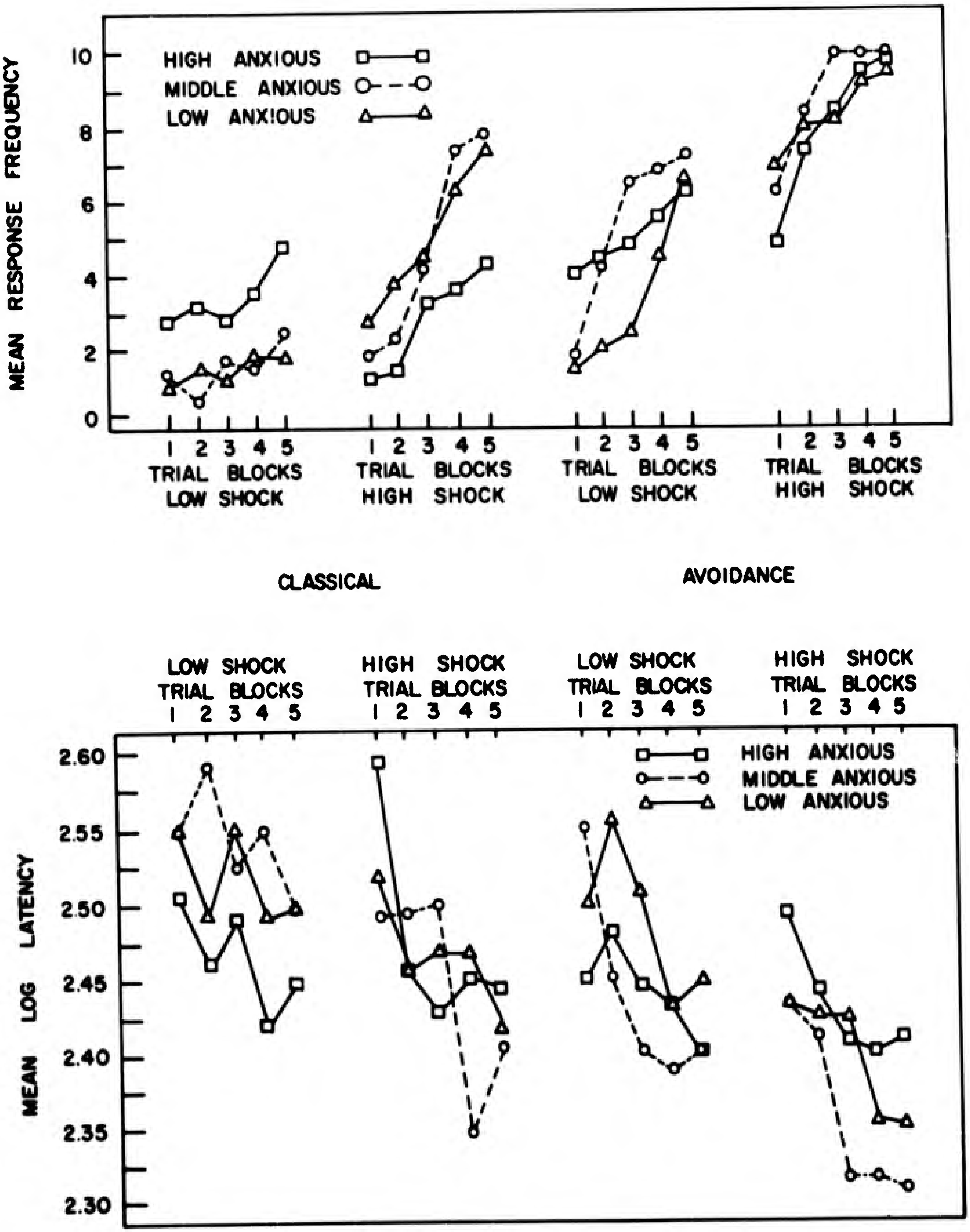
SUMMARIES OF ANALYSES OF VARIANCE OF FREQUENCY AND LATENCY SCORES, STUDY VI

Source of Variation	df	Frequency MS	Frequency F	Latency MS	Latency F
Between Subjects	119				
Conditioning Procedure (C)	1	1,700.16	53.73**	.576	30.16**
Shock (S)	1	1,255.70	39.68**	.430	22.51**
Anxiety (A)	2	13.27	1	.026	1.35
C x S	1	125.12	3.95*	.013	1
C x A	2	18.30	1	.057	2.96
S x A	2	112.36	3.56*	.107	5.59*
C x A x S	2	29.89	1	.016	1
(Subject within grp. Error)	(108)	(31.64)			
Within Subjects	480				
Trial Blocks (Tr)	4	234.18	87.94**	.020	1.83
C x Tr	4	7.54	2.80*	.101	9.09**
S x Tr	4	10.88	4.09**	.046	4.18**
A x Tr	8	28.09	10.52**	.035	3.18**
C x S x Tr	4	26.88	10.09**	.023	2.11*
C x A x Tr	8	4.50	1.69	.020	1.81
S x A x Tr	8	3.34	1.30	.024	2.14**
C x A x S x Tr	8	8.13	3.05**	.060	5.14**
(Tr x Subject within grp. Error)	(432)	(2.6629)		.011	

* $p < .05$
** $p < .01$

Figure 4

Mean Response Frequency and Latency per Trial Block in Each Condition for Each Anxiety Group, Study VI



Anxiety differences did not parallel shock effects; they had no overall effect on either frequency or latency. However, anxiety interacted with other variables in a most important way. There was a shock x anxiety interaction ($p < .05$) in both frequency and latency results (see figure 4). Response frequency was directly related to anxiety level (following expectation) in the low shock condition; in the high schock condition, however, this relationship was generally reversed, with high-anxious subjects giving fewest responses and little difference between low- and middle-anxious subjects. Parallel results were found on the latency criterion, with anxiety related as expected to latency (ie, inversely) under low shock but with a generally opposite relationship under high shock. In each instance, under high shock the relationship of anxiety to performance was slightly curvilinear, with middle-anxious subjects having the greatest frequency and lowest latency of response.

Overall, these results closely parallelled Arousal theory expectations. In fact, subjects of all anxiety groups produced considerably more responses under high than under low shock, and under avoidance than under classical procedures, except that high anxious subjects under classical procedures produced fewer responses under high than under low shock. (It should be noted that this trend, although dramatic, was not supported by a significant conditioning x shock x anxiety interaction).

There were also significant conditioning x shock x anxiety x trials interactions for both frequency and latency, and a significant shock x anxiety x trials interaction for latency. These reflected complex relationships not permitting simple interpretation; they are illustrated in figure 4.

Conclusions

In general, this study supported the use of this as a classical conditioning task. Frequency and latency criteria were moderately associated within groups (enough so to permit combining criteria for item analyses), and both reflected the learning function expected in a conditioning task. The importance of incentive in such a task was dramatically confirmed, and with it the suspicion of any conditioning task with a built-in incentive function of the UCS. However, when the incentive function was absent, in the classical condition, the expected Drive effect of UCS intensity was found.

The results concerning anxiety, however, were far from expected. Anxiety as a Drive measure parallelled UCS intensity perfectly - but only for subjects in the low shock condition. In the high shock condition the relationship was reversed, and was also somewhat curvilinear (with middle-anxiety subjects giving highest frequencies and lowest latencies). This reversal was especially manifest under the classical procedure, in which high-anxious, high shock subjects gave fewer responses than high-anxious, low shock subjects. Subjects in other groups uniformly gave more responses under high than under low shock.

These unexpected results suggest several alternative interpretations. They may reflect the lack of validity of the MAS, although the expected relationship obtained under the low shock condition suggests to the contrary. They may also reflect lack of validity of this task as a classical conditioning task, or at least as one in which Drive effects can be clearly isolated and defined. Finally, however, they may reflect on the relative validity of Drive theory and Arousal theory hypotheses.

In Arousal theory terms these results suggest that (1) Arousal was highest among high anxious subjects under high shock conditions in which the shock could not be avoided and (2) this level of Arousal exceeded the optimum level in this task, thus leading to decrement of performance. As mentioned earlier, Broen and Storms (ref. 20) have proposed a slight amendment to Spence's formulations concerning Drive, whereby the latter could encompass these data (as well as other data consistent with the Arousal theory formulation). Their proposal was that one can assume that E (effective reaction potential) may reach asymptote, perhaps at some physiological limit, beyond which increase in either H or D will not increase E. As D increases beyond this point, it serves only to increase the E's of responses lower in the habit-hierarchy - ie, those still at sub-asymptotic level. As these response tendencies increase, then, they may more frequently reach threshold value. If their occurrence inhibits occurrence of the dominant response, the latter will then occur less frequently: ie, there will be a decrement in performance.

This proposition of Broen and Storms, and the more general proposition of the Arousal theorists, received some support from unsystematic observation of the subjects' behavior. It was apparent that many responses other than finger retraction were often elicited by the UCS: head-jerking, wincing and grimacing, jerking the foot or contralateral arm, etc. These responses also seemed to occur on unreinforced (nonshock) trials; that is, it appeared that they became conditioned to the CS. It also appeared that they occurred more frequently under high shock than low shock, and under classical than under avoidance procedures. The latter two impressions were uncertain, however, since no systematic frequency counts were made. Nevertheless, they suggested that the Broen and Storms proposition may account for the results regarding finger-retraction, particularly those concerning decrement of finger-retractions by high-anxious subjects, in the high shock, classical condition.

The next study was conducted to test this formulation.

STUDY VII

Purpose

The purpose of this study was as explained above. There was no intention of using these results in test development. The procedure required did not permit testing enough subjects for item analysis procedures, as will be apparent below.

In brief, the purpose was to learn whether responses other than finger-retraction become conditioned to the UCS, and, if so, to learn how their occurrence is related to intensity of UCS, conditioning procedure (classical vs avoidance), and anxiety, singly and in combination.

Method

Apparatus

The same apparatus as that used in Study VI was used with the following modifications and additions. A 16 mm motion picture camera with wide angle lens was used for recording all general activity in addition to finger-retraction. The camera, hidden from view, was focused on the subject through a small aperture in a booth which was lined to mask camera noise.

The increased level of background noise caused by the camera equipment necessitated an increase in the masking tone, provided by the audio oscillator, from the 40 db SPL used in Study VI to 50 db SPL. To increase the intensity of the CS correspondingly, a doorbell buzzer of approximately 72 db SPL intensity was substituted for the induction coil (61 db SPL) which generated the CS in Study VI.

Five lights, mounted on the side of the chair, were used as event markers to facilitate film analyses. The subject was asked to sit errect, keep his left hand on his knee, and fixate on a small X placed on the wall directly in front of him. A panel above the lights made it impossible for the subject to see the event-marker lights. The room was sufficiently illuminated that the onset of the lights did not cause a reflection in his line of vision and was not noticed by the subject. Questioning of the subjects following the experiment revealed that they were not conscious of the flashing lights.

Subjects

Subjects were 27 male introductory psychology students stratified into high-, middle-, and low-anxiety groups on the basis of scores obtained on the MAS. Low-anxiety subjects had scores of 15 and below, middle-anxiety subjects had scores from 16 to 20, and high-anxiety, scores of 21 and above. (These cutoff points represent the low, middle, and upper thirds of a distribution of scores obtained from administration of the MAS to 500 introductory psychology students one year prior to this study).

Experimental Design

Except for the use of a control group, the design was the same as that for the first experiment. Half the subjects in each anxiety group were conditioned by the classical procedure and half by the avoidance procedure. Half of the subjects in each conditioning group received high shock, and half received low shock, as the UCS.

Three subjects, one from each anxiety group, were assigned to a control condition. The purpose was to determine the base rate of random movement, unrelated to CS or UCS, which might be erroneously considered a CR or UCR. For controls the procedure was the same as that for subjects in the experimental groups, except that shock was never presented.

The design was again a 2 x 2 x 3 x 5 (conditioning procedure x shock intensity x anxiety x trials) factorial analysis of variance design, with repeated measures over trials. The no-conditioning control group was not included in the analysis of variance. Latency measures were not analyzed. Dependent variables were (1) frequency of finger-retraction and (2) frequency of all other, "irrelevant," responses (see below).

Procedure

The experimental procedure was identical to that of Study VI. The following procedure was employed for tabulating frequency of "irrelevant" response. A transformer was wired to the motion picture projector so that the films could be projected at varying speeds. Each film was viewed several times, at varying speeds, in scoring each subject. The event-marker lights on the subject's chair identified the onset of CS (and UCS) to identify the interval in which a movement could be considered a CR or UCR. The lights also identified the occurrence of nonreinforced test trials, on which the UCS was not presented and responses could be considered CRs.

To determine rater reliability, films of 10 of the 27 subjects were selected randomly and scored independently by the experimenter and an assistant. The assistant was familiar with the conditioning procedure but not with the purposes of the study. Both the experimenter and the other judge recorded, for each test trial for each subject, the irrelevant responses observed. These responses were categorized by muscle group, direction, and amount of movement. For each muscle-group category, a correlation (Pearson $\underline{r}$) was computed between the number of times for each block of trials a response was observed by each rater.

Those response categories for which there was acceptable agreement, using a correlation of 0.90 or above as the criterion, were retained for analysis. Other response categories were either eliminated or combined to obtain acceptable agreement. For example, the response categories leg extension, leg flexion, and leg retraction were combined to make a response category called leg movement. A list of response categories for which there was acceptable agreement (correlation coefficient of 0.90 or above) is available in ref. 15. They were flinching, hunching and rearing, arm movement, hand flexion, leg movement, stepping, and grimacing. Detailed information concerning scoring methods is also available in ref. 15.

Following this preliminary procedure, films for all 27 subjects were scored by the experimenter, using as response criteria only those categories for which there was satisfactory agreement.

Results

There was no single irrelevant response which was emitted by all subjects, nor was any individual subject entirely consistent in emitting any particular irrelevant response. Nevertheless, a variety of irrelevant responses were emitted during unreinforced trials. Two of the three subjects in the no-conditioning control group appeared to emit irrelevant responses, apparently from postural shifts which coincided by chance with the incidence of a "trial." These responses were markedly less frequent than the irrelevant responses on unreinforced trials of experimental subjects. The irrelevant responses of the experimental subjects were not simply spontaneous postural shifts. The type and vigor of the responses (eg, grimacing) of the experimental subjects left little doubt that these responses were more than random.

In analyzing results concerning irrelevant responses, each subject was scored simply for presence or absence of one or more irrelevant responses on each unreinforced trial. Type and magnitude of response was disregarded.

Two initial analyses of variance, one for frequency of finger-retraction and the other for frequency of irrelevant responses, were conducted to determine the effects of conditioning procedure on acquisition of finger-retraction and irrelevant responses (see table VIII). Shock and anxiety were disregarded in these analyses.

As in Study VI, the avoidance (incentive present) procedure increased frequency of finger-retraction responses ($p<.01$). For irrelevant responses the trend, although nonsignificant, was opposite.

In the classical condition irrelevant responses increased regularly across trials, parallel to the increase of finger-retraction responses. In the avoidance condition irrelevant responses increased initially, but then extinguished (conditioning x trials, $p<.05$). This extinction coincided with the point where finger-retraction responses approached 100 per unit frequency (see figure 5). The shock incentive in the avoidance condition served as a selective reinforcer, then, for finger-retraction responses. Without the selective reinforcer response-acquisition was as rapid for irrelevant as for finger-retraction responses.

Summing across conditioning procedures, there was an overall increase across trials in both finger-retraction and irrelevant responses (trials main effect, $p<.01$ in each case). For irrelevant responses, however, this was attributable entirely to response increase in the classical condition, as is apparent in figure 5.

Separate analyses were then conducted (table VIII) to determine the effects of shock and anxiety in each conditioning procedure.

In the classical condition both kinds of responses increased across trials ($p<.01$), and increased more rapidly ($p<.05$ and $p<.01$) under high than under low shock. The frequency of the two kinds of responses, under each shock level and on each trial block, was strikingly similar (see figure 5). There were no differences among subjects of different anxiety levels.

In the avoidance condition, finger-retraction responses increased rapidly ($p<.01$) under both shock levels; anxiety had no significant effect. For high-anxious subjects irrelevant responses also increased across trials; but low-anxious subjects emitted progressively fewer irrelevant responses, and middle-anxious subjects showed no regular increase or decrease (anxiety x trials, $p<.05$).

TABLE VIII

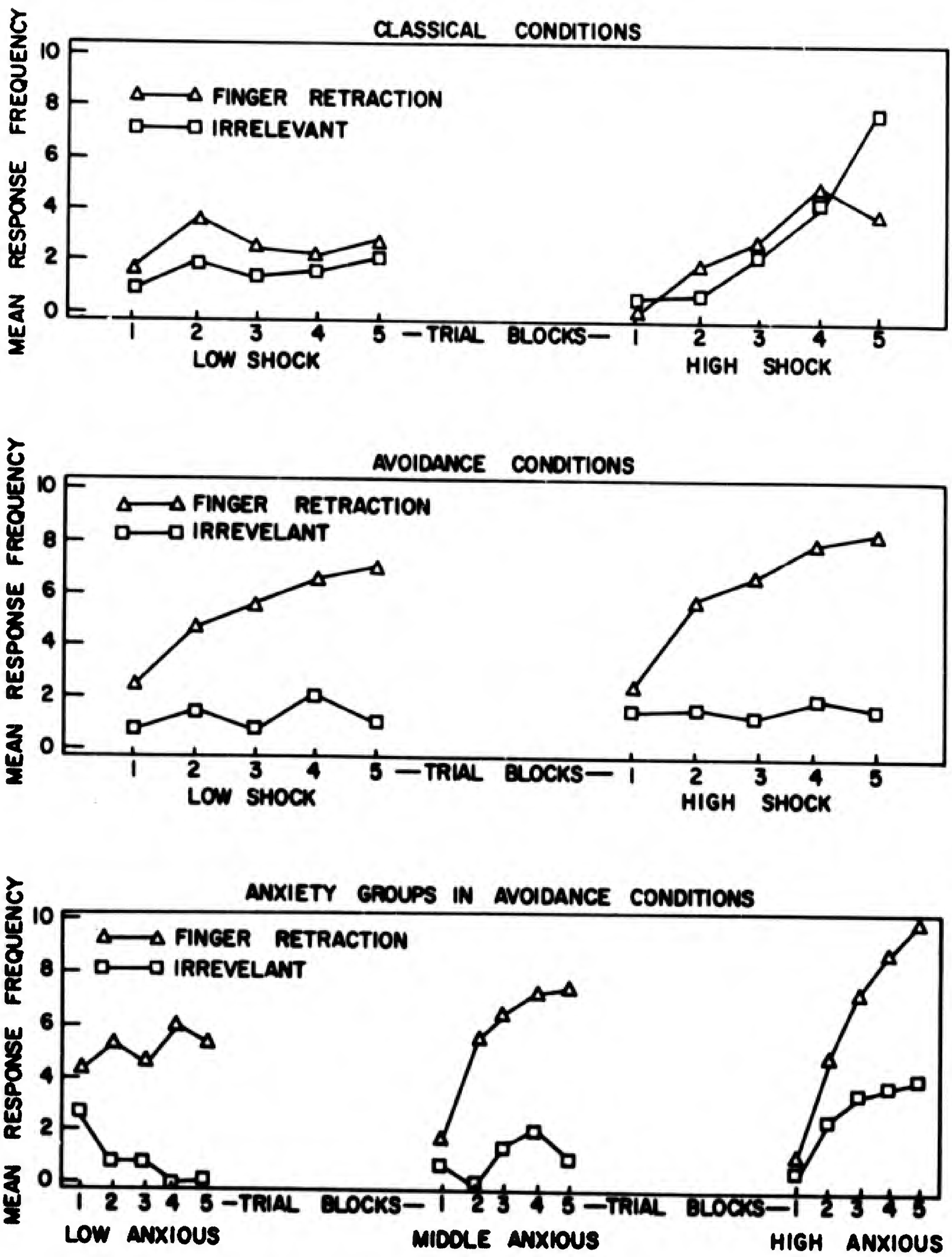
SUMMARY OF ANALYSES OF VARIANCE OF FREQUENCY OF FINGER-RETRACTION AND IRRELEVANT RESPONSES, STUDY VII

		Preliminary Analysis					Classical				Avoidance			
		Fing. Ret.		Irr. Resp.			Fing. Ret.		Irr. Resp.		Fing. Ret.		Irr. Resp.	
	df-	MS	F	MS	F	df	MS	F	MS	F	MS	F	MS	F
Between Subjects	23					35								
Conditioning Procedure (C)	1	313.6	8.2**	23.4	1	-	-	-	-	-	-	-	-	-
Shock (S)	-	-	-	-	-	1	.1	1	12.8	1	4.0	1	2.5	1
Anxiety (A)	-	-	-	-	-	2	12.5	1	4.0	1	3.4	1	7.7	1
S x A	-	-	-	-	-	2	2.9	1	7.9	1	28.9	1	7.6	1
Subj. within grp. Error)	(22)	(38.2)		(24.9)		(30)	(16.5)		(15.7)		(61.5)		(10.3)	1
Within Subjects	96					144								
Trial Blocks (Tr)	4	61.0	15.0**	21.1	4.6**	4	4.3	3.3*	11.5	4.7**	19.1	7.5**	.7	1
C x Tr	4	6.7	1.6	15.7	3.4*	-	-	-	-	-	-	-	-	-
S x Tr	-	-	-	-	-	4	3.4	2.6*	8.6	3.5**	.4	1	.4	2.1*
A x Tr	-	-	-	-	-	8	1.2	1	1.1	1	3.5	1.4	2.2	1
S x A x Tr	-	-	-	-	-	8	.8	1	.5	1	1.8	1	.8	1
(Tr x Subj. within grp. Error)	(88)	(4.1)				(120)	(1.3)		(2.5)		(2.6)		(1.1)	

* p<.05
** p<.01

Figure 5

Mean Frequency per Trial Block of Finger-Retraction and Irrelevant Responses, Study VII



In summary, all subjects acquired both finger-retraction and irrelevant responses in the classical condition, and acquired them more rapidly under high than under low shock. In the avoidance condition all subjects acquired the finger-retraction response, and this response was again more frequent in avoidance than in classical conditions. High-anxious subjects also acquired irrelevant responses in the avoidance condition, but these responses extinguished for low-anxious subjects and were not stable for middle-anxious subjects.

Conclusions

The shock x anxiety interaction for finger-retraction responses, found in Study VI, was nonsignificant in this study, as were several other previously significant effects. The greatly restricted number of subjects in the present study may have obscured such effects, simply through loss of statistical power. In any case, the results of this study did clarify the findings of Study VI.

In the classical conditioning procedure, responses other than finger-retraction were conditioned to the CS. These responses were not incompatible with finger-retraction responses; both kinds of responses often occurred on the same trial. Nevertheless they may have competed with finger-retraction responses in rate of acquisition. An incentive for the finger-retraction response, as in the avoidance procedure, serves as a selective reinforcer. With it, finger-retraction responses were acquired more rapidly, and other responses generally (if nonsignificantly) occurred less often. Without the selective incentive, both kinds of responses were acquired at about the same rate, neither one as fast as the finger-retraction response when it had an incentive-attainment function.

In Study VI there was a performance decrement in finger-retraction responses among high-anxious subjects in the high shock, classical condition. This phenomenon seemed more consistent with Arousal theory than with Drive theory predictions. High-anxious subjects may have been performing at higher than optimal Arousal level, even for this very simple task. If one can assume that the effects of anxiety and shock are additive, this proposition is tenable. Observation of the subjects in this condition supported this assumption: the stress level seemed barely tolerable for many subjects, and in fact intolerable for a few.

Broen and Storms' (ref. 20) hypothesis suggested an alternative explanation. Other responses might have been competing with the finger-retraction response at this Drive level, simply because the reaction potential for finger-retraction had reached asymptote. This was not directly supported in this study. Finger-retraction responses never approached 100 per cent frequency, in the classical procedure, as they did in the avoidance procedure. It is difficult to assume, then, that they had approached a physiological limit or asymptote. Also, other responses were not incompatible with the finger-retraction response.

Still, the presence of conditioned responses other than finger-retraction, confirmed in this study, might affect rate of finger-retraction responses; and high-anxious subjects acquired irrelevant responses even under avoidance conditions, whereas other subjects did not. The reaction-time studies suggested that high-anxious subjects may tend to have a generalized response-inhibition tendency, when there is available a response which can lead to incentive-attainment. The present results suggest further that these subjects may have a greater tendency toward generalized nonfunctional responses in a stress situation. Neither of these tendencies is foreign to clinical suppositions regarding the behavior of high-anxious subjects. Both, however, would represent contamination of the MAS as a measure of generalized Drive.

As a final point it should be noted that irrelevant responses might in fact serve some incentive function in the classical procedure, as might finger-retraction responses. Neither succeeds in avoiding or escaping the shock. Either, however, might attenuate the unpleasantness of the shock. Any sudden contraction of musculature might reduce the perceived intensity of a simultaneous noxious stimulus. This possibility would be difficult to test. However, the pervasive effects of incentives, found in both reaction-time and finger-retraction studies, imply that the <u>absence</u> of incentive-mediated phenomena may be more unlikely than their presence.

STUDY VIII

Purpose

Study VIII, like Study IV, was conducted for the purpose of acquiring criterion data for test development. Only secondarily, it was hoped that this study could provide replication of some aspects of Study VI, the first finger-retraction study.

All subjects were tested under the low shock, classical condition. This was selected because (1) it seemed as "pure" a classical conditioning task as possible, (2) the high shock classical condition was extremely stressful to the subjects, (3) the intensity of the noxious UCS in the low shock condition was also more likely to be representative of laboratory tasks commonly used, and (4) it was in the low shock condition particularly that, in Study V, the predicted relationship of individual differences in Drive to performance was obtained.

In Study VI, high-anxious subjects had shown superior performance in this condition, but there were no differences between low-anxious and middle-anxious subjects. This might simply suggest that the relation of MAS scores to individual differences in Drive is nonlinear. A difference of five points in scores at the lower end of the scale might reflect inconsequential differences in Drive, but at the higher end of the scale this difference might reflect substantial Drive differences. The assumption of interval scaling of the MAS may, then, not be tenable. In almost all prior experimental research with the MAS only extreme high-scorers and extreme low-scorers have been used, so this phenomenon could easily have escaped attention.

MAS scores had been found consistently correlated with scores on the M-C SD, which was originally intended strictly as a measure of individual differences in social desirability response-set. This would suggest that the low-scorers on the MAS are particularly biased toward social desirability response-set, and consequently that their scores on the MAS are distorted. The low-anxious group may have included several "true" high-anxious and middle-anxious subjects.

Consequently, subjects were doubly categorized, on both the MAS and the M-C SD, as separate factors in analyses of variance. It was expected that the relationship between MAS scores and performance might not be apparent for high M-C SD scorers, but might be more clear for other subjects. Subjects were also divided into five, rather than three, categories based on MAS scores, to explore possible true curvilinear relationships of anxiety to performance. By using the M-C SD for independent categorization, it was also possible to explore relationships of need for approval (or social desirability response-set) to performance. There was little or no reason for expecting such a relationship, except for the consistency with which M-C SD scores had been found to be related to performance on other kinds of tasks.

Method

Apparatus

The same apparatus, sans motion picture equipment, was employed as in Study VI and Study VII, with one exception. A slightly altered, more sensitive microswitch was employed. This was substituted for the original microswitch for the purpose of recording responses which might otherwise be of insufficient magnitude to be recorded.

Subjects

As in Study IV, subjects were recruited from two freshman AFROTC sections; total number of subjects was 107. All subjects had been tested previously on the MAS, M-C SD, LC, EPPS, ACL, EIS, and IS.

Experimental Design

All subjects were tested under the same (low shock, classical) conditions, for a total of 50 test trials (five blocks) randomly distributed among 200 reinforced trials. For analysis of variance, subjects were doubly categorized, into five categories based on MAS scores and into three categories based on M-C SD scores. The result was a 5 x 3 x 5 (anxiety x M-C SD x trial blocks) factorial analysis of variance design, with repeated measures over trials. An unweighted means analysis was necessary, since the negative correlation of MAS and M-C SD scores led to under-representation of high-anxiety, high M-C SD and low-anxiety, low M-C SD groups.

The dependent variable was frequency of CR. The latency criterion was not employed because it seemed generally less sensitive than the frequency criterion in Study VI.

The criterion for defining a conditioned response was slightly altered from that used in the previous two studies. Any response occurring on a test (no UCS) trial within 610 ms was considered a CR. The previous two studies had employed an avoidance, as well as a classical, condition. In the avoidance condition a response on a nontest trial of more than 400 ms did not succeed in avoiding the UCS. On test trials a response was considered a CR only if it occurred within this 400 ms interval, plus 10 ms to include responses initiated before the end of the CS-UCS interval. The same criterion was used to define a CR in the classical condition, since a major interest was comparing avoidance vs classical procedures with other variables held constant.

In addition, in this study a response on a non-test trial was considered a CR if it occurred within 500 ms of CS onset. From inspection of raw data it was apparent that the latency of the UCR to the UCS was greater than 100 ms. Adding these responses, which appeared to be bona fide conditioned responses, provided a more sensitive dependent variable measure, since it in effect increased the number of trials on which a CR might occur to 50 per block. The number of CRs as defined by this more liberal criterion correlated 0.72 with the number of CRs on test trials alone, so the altered criterion seemed unlikely to change results except for the increase of sensitivity and statistical power.

Procedure

The procedure was unchanged from that of Study VI and Study VII.

Results

Several subjects had to be discarded because they failed to give a UCR to any shock within the predefined, low-shock range. Equipment failure resulted in discarding several others. A total of 85 subjects were retained for data analysis.

The distribution of frequency scores for these subjects was markedly skewed. Many subjects emitted no CRs at all, even though they regularly emitted UCRs. Over 90 per cent emitted fewer than 12 CRs out of a possible 50 on test trials; the remaining subjects ranged from 20 to 50 CRs. A log transformation was performed to normalize scores for analysis of variance.

Response frequency significantly ($p < .01$) decreased across trials. After an early, slight acquisition of the finger-retraction response, response frequency followed an extinction, rather than acquisition, curve. This was apparent also when response frequency was defined strictly in terms of responses on unreinforced trials, even if these included only responses of 410 ms latency; various definitions of conditioned responses yielded identical results.

Anxiety differences were unrelated to response frequency; the trend was both negligible and inconsistent. This was true of MAS groups in all three M-C SD categories.

Response frequency was related, however, to M-C SD groupings ($p < .05$). The three M-C SD groups had equivalent frequencies on the first trial block; thereafter, response frequency decreased rapidly for low M-C SD subjects, and less rapidly for high- and middle-scorers, in that order (M-C SD x Tr, $p < .05$). There was, then, a curvilinear relation of M-C SD scores to response frequency.

Conclusions

It was obvious that the task failed as a classical conditioning task in this study. Consequently, the study could not be considered a replication of Study VI.

TABLE IX

SUMMARY OF ANALYSIS OF VARIANCE OF LOG RESPONSE FREQUENCY SCORES, STUDY VIII

Source of Variation	df	MS	F
Between Subjects	84		
Anxiety (A)	4	.056	1
M-C SD (M)	2	.678	3.30*
AM	8	.212	1.03
(Subj. within grp. Error)	(70)	(.205)	
Within Subjects	340		
Trial Blocks (Tr)	4	.375	8.75**
A x Tr	16	.032	1
M x Tr	8	.090	2.10*
A x M x Tr	32	.049	1.15
(Tr x Subj. within grp. Error)	(280)	(.043)	

* $(p < .05)$
** $(p < .01)$

The reasons for the failure were not obvious, however. It was fully ascertained that there was no apparatus failure, including the apparatus for generating and delivering shock. All subjects reported that they felt the shock; and calibration was rechecked and found still accurate. Even the instructions were identical to those used before. The only obvious change from studies VI and VII was that there was a different experimenter. The experimenter in this study had assisted in the previous studies, however, and had been present as the motion picture camera operator in Study VII. He was fully familiar with the procedure and duplicated it without apparent change.

Only one explanation of these results can be suggested. It appeared that many subjects may have been, purposely or unconsciously, pushing down on the microswitch, and consequently inhibiting the finger-retraction response. A strong effort was made to discourage subjects from doing this, short of telling them to purposely retract their fingers. This effort seemed often to be unsuccessful, although objective evidence was minimal.

It is possible that personality differences between experimenters, or differences in inflection while delivering instructions, were related to this phenomenon. Several times in the present study when a subject gave a CR he seemed embarrassed, and sometimes failed even to give UCRs on the next few trials even though he was assured it was "OK" to give a CR.

In the present study the experimenter recorded the subject's estimate of the number of buzzes (CSs) there had been. This was not done in studies VI and VII. This may have been responsible for some minimal cue that buzzer-counting was more important than it appeared to subjects in the other studies. Subjects in the present study gave much more accurate counts, in fact, than the few in previous studies who were asked to give a count. It is possible that subjects in this study adopted a more attentive posture to keep closer count, and that this included a press-down response. Again, unsystematic observation suggested that this was sometimes true.

Response inhibition may, then, have contaminated results, as it seemed to in the reaction-time studies. And again, incentive variables may have been at work. Embarrassment at the occurrence of a CR suggests fear of the disapproval of the experimenter; and, in fact, the high need for approval subjects gave fewest CRs, although middle-scoring subjects gave the most responses. Achievement (or approval) incentives associated with buzzer-counting may have contributed to these results. Some subjects also seemed to consider it a challenge to manliness to resist the shock.

With these questions, the data from Study VI were reinspected. A few subjects in that study were discarded from the low shock classical condition (and from the entire study) because of failure to give a UCR to the shock. Several others in this condition also failed to give CRs, even though they gave UCRs. In these respects the results of the two studies were quite similar.

One other phenomenon was observed in the Study VI results which had been overlooked previously. When the performance of each subject was examined individually, there was usually some point (trial block) at which the subject's performance across trials was markedly discontinuous. Response frequency often went from around 30 per cent to 90 or 100 per cent from one block to the next. When this occurred there was also a sharp drop in latency (ie, latency on only those trials in which a CR did occur). This point occurred at different times for different subjects. When it did occur, performance almost always stayed at or above this high level. This occurred earlier in performance under avoidance than under classical conditions, and under high than under low shock.

The consistent discontinuity phenomena raised suspicion that subjects in the earlier studies may have, at different points, begun to respond more or less voluntarily. In turn, this suggested, as was considered earlier, that the finger-retraction response may serve to attenuate the unpleasantness of the shock. A press-down response of the finger might accomplish the same function. Incentive phenomena may again have contaminated results of all three finger-retraction studies, even when the results seemed to reflect Drive phenomena (in Study VI under classical conditions).

SUMMARY: MOTIVATION AND FINGER-RETRACTION CONDITIONING

Motivational Variables in Finger-Retraction Conditioning

Once more, it is moot whether Drive effects were in fact observed in these studies. The evidence of incentive effects was inferential. Nevertheless, results were not consistent with reasonable expectations of Drive effects; and incentive effects seem increasingly omnipresent.

Further effort might still succeed in instrumenting the finger-retraction task in a way to elucidate Drive phenomena. Two alterations can be suggested.

The press-down response, with associated incentive effects, may be a major contaminant. To avoid this, a different procedure for recording responses might be devised, which would not require the subject to support his finger on a microswitch. EMG recordings might be taken from the musculature involved in the finger-retraction response. The subject's hand and forearm could be allowed to hang, unsupported, over a chair arm. With no supporting surface, anticipatory or inhibitory responses might be less likely. EMG recordings could in fact detect movement of the hand in any way or direction. Any movement of the hand, coincident with CS, could equally well be considered a CR.

Second, voluntary responses, including the press-down response, might be partly obviated if the conditioning took place in conditions in which the subject was more fully distracted. A demanding intellectual task unrelated to CS or UCS might accomplish this.

These seem, frankly, to be remote hopes. However, a successfully instrumented classical conditioning task would be of great value, particularly in investigating Drive and stress phenomena. Eyeblink conditioning appears even more to be complicated by incentive variables; and cardiac and GSR conditioning are still beset by problems. Possibly a different UCR, such as the pupillary reflex, could be employed. And possibly further attempts to alter the finger-retraction task will be successful.

Implications for Test Construction

Study VIII clearly demonstrated that this task is of limited value for developing tests of individual differences in Drive. It may, of course, be useful in developing tests of incentive-relevant variables. Incentives relevant to performance on this task were not clearly identified, however. It remains possible that the results of Study VI do reflect Drive differences, especially among subjects in the low shock classical condition. These data were treated as the most likely to be criteria of Drive differences in test development. Confidence must be limited in the result, however.

SECTION VI
MOTIVATIONAL CORRELATES OF VERBAL LEARNING

RATIONALE AND METHOD FOR VERBAL LEARNING TASK

Rationale

The previous studies confirmed the importance of motivational variables in performance of simple motor tasks. It was also apparent that these variables are complex, and that their relation to performance is complexly determined by many other variables.

Verbal learning tasks might well be susceptible to different motivational effects, however, than are simple motor tasks. Verbal tasks are more complex, and more clearly "social" in ramifications, than motor tasks. Different incentive, higher-order habit, and cue variables may be predominant in verbal tasks. Drive and arousal phenomena may be even more obscured in these tasks than in motor tasks. Consequently, motivational tests which predict performance on motor tasks may be virtually useless with verbal tasks.

This phase of the research was directed toward investigation of the commonalities and differences, between simple motor and verbal tasks, in motivational phenomena.

To maintain as much continuity as possible, a verbal learning task was selected which, as much as possible, shared characteristics with the simple motor tasks employed for test development. Because of the particular interest in Drive phenomena, one task variable of primary interest was dominance of correct response. It was particularly desirable, then, to select a verbal task in which the dominance of the correct response could be systematically manipulated in some way, as in the reaction-time tasks.

The verbal paired-associates task previously employed by Chiles (ref. 39) was selected for this purpose. This task was designated to permit exploration of differential effects of Drive on performance when the correct response is dominant and when it is not. It had provided results (ref. 39) conforming generally to expectations of Drive theory.

Both Drive and social incentive effects were investigated, while data were being collected for test development. Drive was experimentally manipulated, as in the Chiles procedure, by manipulation of a noxious UCS. MAS scores were also used, as in Ramond's study (ref. 40), as a Drive measure. An exactly parallel procedure was adopted for exploring social incentive effects within the same studies.

Rather than extend the research to other social incentives (eg, n Ach), the primary interest was in need for approval. This variable had already been found to be important even in the reaction-time tasks, as in many other tasks employed by other investigators. The M-C SD had demonstrated far greater power and validity than had any nAch test. Also, considerable data had been collected with the M-C SD in the earlier studies and concentrating on this variable provided greater continuity and coherence in the total data-collection efforts.

Research concerning need for approval had not clarified whether this variable pertains to the positive incentive of approval or to the negative incentive of disapproval, or to both. This distinction, although admittedly crude, might be important. A procedure was adopted which permitted some exploration of this question. Need for approval was both experimentally manipulated and measured by M-C SD.

Method for Verbal Learning Task

Verbal Learning Task

The task was similar to that used by Chiles (ref. 39), adapted to fit a 50-min experimental session for administration and to use available apparatus. The stimuli were 16 triads of words. A stimulus word was on the left and two response words, one under the other, were on the right. All words were two-syllable adjectives. The subject's task was to learn which of the two response words was correct for each stimulus word.

In each of the triads, one of the response words was synonymous to the stimulus word and the other was unrelated to the stimulus word. The former was assumed to have higher associative value to the stimulus word than the latter. The synonym was assumed, then, to be the dominant response. For 8 of the 16 triads, the dominant response word was correct. For the other eight triads the unrelated stimulus word was correct.

Each pair of response words appeared with two different stimulus words. Each was the dominant response alternative with one of the stimuli, and each was the correct response in one of the two pairings. There was a random order of triads in which the dominant response was correct and those in which the unrelated response was correct.

The triads were arranged in four different random orders. Each resultant order was denoted a "list." The four lists were presented sequentially, always in the same sequence, for six total presentations of the four lists. There were, then, 24 trials on each triad. The triads, in the four different orders, are available in ref. 41.

The practice list consisted of 16 triads of unrelated two-syllable nouns. It was presented for six trials, with the triads always in the same order (see ref. 41).

Distractor Task

The Drive stimulus (shock) was delivered in conjunction with a distractor task to make the shock UCS irrelevant to verbal learning performance. The subject was told that the study was concerned with the development of Air Force high altitude warning systems, and that the specific purpose of the study was to "determine whether a pilot would respond more quickly to a buzzer or electric shock." Because pilots are kept busy operating their aircraft, "serving a basically mental function," the subject was to be'kept busy" with a "basically mental activity" (the verbal learning task). Otherwise the instructions were essentially those of Chiles (ref. 39). The subject was told to turn off the shock (or buzzer, depending on the experimental condition) as quickly as possible, and was told the experimenter would say "right" when the subject gave the correct response word during the verbal learning task but would say nothing when the subject gave the wrong response.

Shock conditions

A switch, which would break either a buzzer or a shock locking circuit, was located on a table to the subject's right as a response key for the distractor task. The circuits were locked by switches controlled by the experimenter. The shock, for subjects in the shock group, was administered to the subject's right calf by electrodes approximately 4 1/2 in. apart. The shock level which continued to elicit a slight twitch or grimace was employed for the Drive stimulus. This level was determined prior to testing, and readjusted as necessary during testing. The buzzer, for subjects in the nonshock group, delivered an auditory stimulus of approximately 66 db.

The shock (or buzzer) was activated at the end of each trial (ie, complete list of 16 triads) and persisted until the subject depressed the response switch. Since no shock (or buzzer) was delivered after the last trial, there were 23 presentations of the shock (or buzz).

Approval Conditions

Three approval conditions were employed: neutral, approving, and disapproving. The latter two were intended to alter the expectancy of, respectively, attaining the experimenter's approval and avoiding his disapproval.

In the approving condition, the experimenter behaved in a generally cordial, friendly, and approving manner, and expressed friendly interest in the subject (his background, scholastic status, etc) before the task began. After the eighth and twelfth trials the experimenter praised and thanked the subject for his effort and attitude. Here, as in the disapproving condition, care was taken to relate the approving and disapproving comments to "effort" and "attitude," rather than to performance per se. This was done to minimize the extent to which approving comments might be perceived as relevant to achievement incentives, rather than to approval incentives as such. In the disapproving condition subjects sometimes asked, in effect, what "they were doing wrong" - ie, whether they were being criticized for their verbal learning performance or for their reaction-time performance. When this occurred, the experimenter dwelt on "general attitude and effort."

In the disapproving condition, the experimenter was generally aloof, impersonal, and gruff, and called the subject by his last name. The general impression attempted was one of unfriendliness, grouchiness, or being in a bad mood. After the eighth and twelfth trials the experimenter severely criticized the subject for his poor attitude and lack of effort. This criticism appeared to be clearly disturbing to the subject; as mentioned, the subject often tried to clarify what he was being criticized for, and seemed genuinely upset. At the end of testing, the experimenter attempted to remove any possible adverse effect of this procedure by explaining that he was in a bad mood that day because "everything had gone wrong" and apologizing to the subject for "taking it out" on him, all the while behaving in a friendly and apologetic manner and assuring the subject that he had "really done quite well." This seemed ultimately to reassure the subject.

In the neutral condition, the experimenter made neither approving nor disapproving comments, and generally behaved in a businesslike but not unfriendly manner. The attempt was to replicate as faithfully as possible the usual behavior of an experimenter in such an experiment.

STUDY IX

Purpose

The primary purpose of this study was simply to collect criterion data for test development. The secondary purpose was to replicate the Ramond (ref. 40) and Chiles (ref. 39) studies, employing both experimentally-manipulated and MAS measured Drive as independent variables simultaneously The third purpose was to extend investigation of social incentive variables, specifically need for approval. As with Drive, need for approval was both inferred from test scores (the M-C SD) and mediated by experimentally manipulated conditions. The final purpose was to investigate whether M-C SD scores reflect individual differences in need for approval or in need to avoid disapproval, by comparing their effects with those of manipulated approval and disapproval incentives.

Method

Apparatus

The shock and buzzer apparatus was described above, as were the stimulus materials. For presenting the stimulus materials, a slide projector was employed. Each triad appeared on a separate slide. A piece of white paper, 8 1/2 by 11 in. mounted on the wall, served as the projection screen. The projection equipment, as well as apparatus for activating the shock or buzzer UCS, were located behind the subject, as was the experimenter.

The automatic timer with which the projector was equipped permitted a 2.4 sec exposure time for each slide, with minimal variation in exposure time, and a 1.1 sec interval between slides. Interlist and intertrial intervals depended on the time to administer the shock (or buzzer) stimulus and change slide trays in preparation for presenting the next list. The median interlist and intertrial interval was approximately 15 sec, except at the times when experimenter approval or disapproval was delivered.

Subjects

All subjects in this study had previously participated in Study VI, and had been previously tested on the MAS, M-C SD, LC, EPPS, ACL, ISB, and specially devised sentence completion tests in their sophomore AFROTC classes. Of the 108 subjects tested, equipment failure necessitated discarding the data for 13.

Experimental Design

Subjects were randomly assigned to Drive conditions (shock or buzzer) and to approval conditions (approving, disapproving, and neutral). All were tested for 24 trials, 16 items per trial. The dominant response was correct on half the items, and on the other half the nondominant response was correct. Following testing of all subjects, subjects within each of the six experimental treatment conditions were categorized as high, middle, and low-anxious on the basis of MAS scores, using as cutoff points the scores which trichotomized the total population; this categorization represented the anxiety factor. The result was a 2 x 3 x 3 x 2 x 24 (Drive x approval conditions x MAS score x response dominance x trials) factorial analysis of variance (unweighted means), with repeated measures over response dominance and trials. Following this analysis, subjects were repooled within treatment groups and categorized on M-C SD scores, and the analysis repeated substituting M-C SD test score for the MAS test score factor. This was repeated with the LC (locus of Control scale).

The dependent variable was number of correct responses (out of 8) on each trial, computed separately for triads in which the dominant response was correct and those on which it was not.

Procedure

The procedure was as described above.

Results

Summaries of the three analyses of variance are in table X. Since the same subjects appear in the three analyses, the analyses are redundant except for main effects and interactions involving test-score groupings.

The results of the Chiles and Ramond studies were not replicated. Shock did not increase performance when the dominant response word was correct, nor decrease performance when the unrelated response word was correct. Neither shock nor anxiety affected performance overall, or in interaction with response dominance, at any stage of learning. Results were inspected separately for each approval condition; in none were the expected trends apparent. The relation of anxiety to performance was inspected in each of the six separate experimental conditions. In none was there replication of Ramond's findings.

Performance was better when the dominant response was correct than when it was not (response dominance main effect, $p<.01$). This supported the assumption that the synonymous response word was more strongly associated with the stimulus word than was the unrelated response word.

Although approval conditions also failed to have significant overall effects, they did affect performance in interaction with other variables. During later trials, in fact, performance was best in the approving condition and worst in the disapproving condition (approval conditions x trials, $p<.05$).

Two experimental conditions produced markedly poorer performance than did the other conditions: the shock, neutral and no-shock, disapproving conditions (shock x approval conditions, $p<.01$; see figure 6). Shock inhibited performance in the neutral condition but increased performance in the disapproving condition; the approving condition produced high performance both with shock and without shock. Approval conditions had different effects on different subjects. High need for approval subjects performed best under disapproving and worst under neutral conditions; subjects in the low and middle need for approval groups, on the other hand, performed worst under disapproving conditions (approval conditions x M-C SD test scores, $p<.05$). This is illustrated in figure 6. Disapproval also produced poorest performance for external-control and best performance for internal-control subjects (approval conditions x LC test scores, $p<.01$;

see figure 6). After trial 8, when the first explicit disapproving comments were made, high-anxious subjects also performed worst in the disapproving condition (approval conditions x MAS test scores x trials, $p < .05$). Disapproval had little effect on low- or middle-anxious subjects, and the three anxiety groups had roughly equivalent performance under the other approval conditions, at all stages of performance.

The effects of approval conditions on different subjects were manifest almost entirely on items in which the correct response was not the dominant response (approval conditions x M-C SD test score x response dominance, $p < .05$, and approval conditions x LC test scores x response dominance, $p < .01$). There was little effect on performance on items in which the dominant response was correct.

The response dominance x trials interaction ($p < .01$) reflected simply that performance was better, in early trials, on dominant-response-correct than on dominant-response-incorrect items, but that this difference decreased across trials.

Conclusions

The present study failed to replicate the findings of either Ramond or Chiles. Neither shock nor MAS scores were directly related to performance, nor did either significantly interact with response dominance on early trials as in the Ramond and Chiles studies. The results suggested, however, that incentive features and cue functions of various aspects of the experimental situation have marked effects on performance. These may have operated to obscure Drive effects.

The effects of shock on subjects in the various approval conditions indicated that, in the absence of approval comments, shock is a powerful incentive for the distractor reaction-time task. With a neutral experimenter shock depressed verbal learning performance, presumably by enhancing the "distraction" effect of the reaction-time task. Without shock incentive for the reaction-time task, subjects in the neutral approval condition excelled on the verbal task, presumably because it was more interesting than the reaction-time task. The failure of shock to affect performance in the approving condition suggests that the incentive of experimenter approval was effective in influencing subjects to try to do well on both tasks.

TABLE X

SUMMARIES OF ANALYSES OF VARIANCE OF NUMBER OF CORRECT RESPONSES, SUBJECTS GROUPED BY MAS, M-C SD, AND LC SCORES, STUDY IX[1]

Source of Variation	df	MAS		M-C SD		LC	
		MS	F	MS	F	MS	F
Between Subjects	94						
Shock (S)	1	.0					
Approval Conditions (Ap)	2	7.4	1.1				
Test Score (T)	2	5.3	1	3.4	1	2.1	1
S x Ap	2	38.6	5.9**				
S x T	2	1.2	1	6.6	1	5.0	1
Ap x T	4	16.0	2.5	23.0	3.5*	25.4	3.9**
S x Ap x T	4	4.7	1	3.8	1	14.4	2.2
(Subj. within grp. Error)	(77)	(6.5)		(6.6)		(6.5)	
Within Subjects	4465						
Response Dominance (Dom)	1	134.6	55.4**				
S x Dom	1	.0	1				
Ap x Dom	2	2.5	1.0				
T x Dom	2	4.0	1.6	.5	1	.9	1
S x Ap x Dom	2	1.4	1				
S x T x Dom	2	4.3	1.8	2.3	1	1.7	1
Ap x T x Dom	4	2.3	1	8.5	3.3*	13.4	4.9**
S x Ap x T x Dom	4	3.1	1.2	2.1	1	1.3	1
Trials (Tr)	23	31.8	89.8**				
S x Tr	23	.6	1.1				
Ap x Tr	46	.5	1.5*				
T x Tr	46	.2	1	.3	1	.4	1
Dom x Tr	23	2.8	9.6**				
S x Ap x Tr	46	.5	1.4				
S x T x Tr	46	.3	1	.2	1	.2	1
S x Dom x Tr	23	.3	1				
Ap x T x Tr	92	.5	1.4*	.3	1	.5	1.3
Ap x Dom x Tr	46	.2	1				
T x Dom x Tr	46	.3	1	.2	1	.3	1
S x Ap x T x Tr	92	.3	1	.3	1	.5	1.1
S x Ap x Dom x Tr	46	.4	1.2				
S x T x Dom x Tr	46	.2	1	.4	1.2	.2	1
Ap x T x Dom x Tr	92	.4	1.3	.3	1.0	.4	1.1
S x Ap x T x Dom x Tr	92	.3	1.0	.3	1	.3	1

Source of Variation	df	MAS MS	MAS F	M-C SD MS	M-C SD F	LC MS	LC F
(Dom x subj. within grp. Error)	(77)	(2.4)		(2.6)		(2.7)	
(Tr x subj. within grp. Error)	(1771)	(.35)		(.38)		(.40)	
(Dom x Tr x subj. within grp. Error)	(1771)	(.30)		(.32)		(.35)	

* p<.05
** p<.01

[1]The same subjects grouped by test scores on MAS, M-C SD, and LC, respectively. The three analyses are redundant except for main effects and interactions involving test-score groupings. Minor variations in redundant MS terms are attributable to artifacts of the unweighted means analysis of variance procedure.

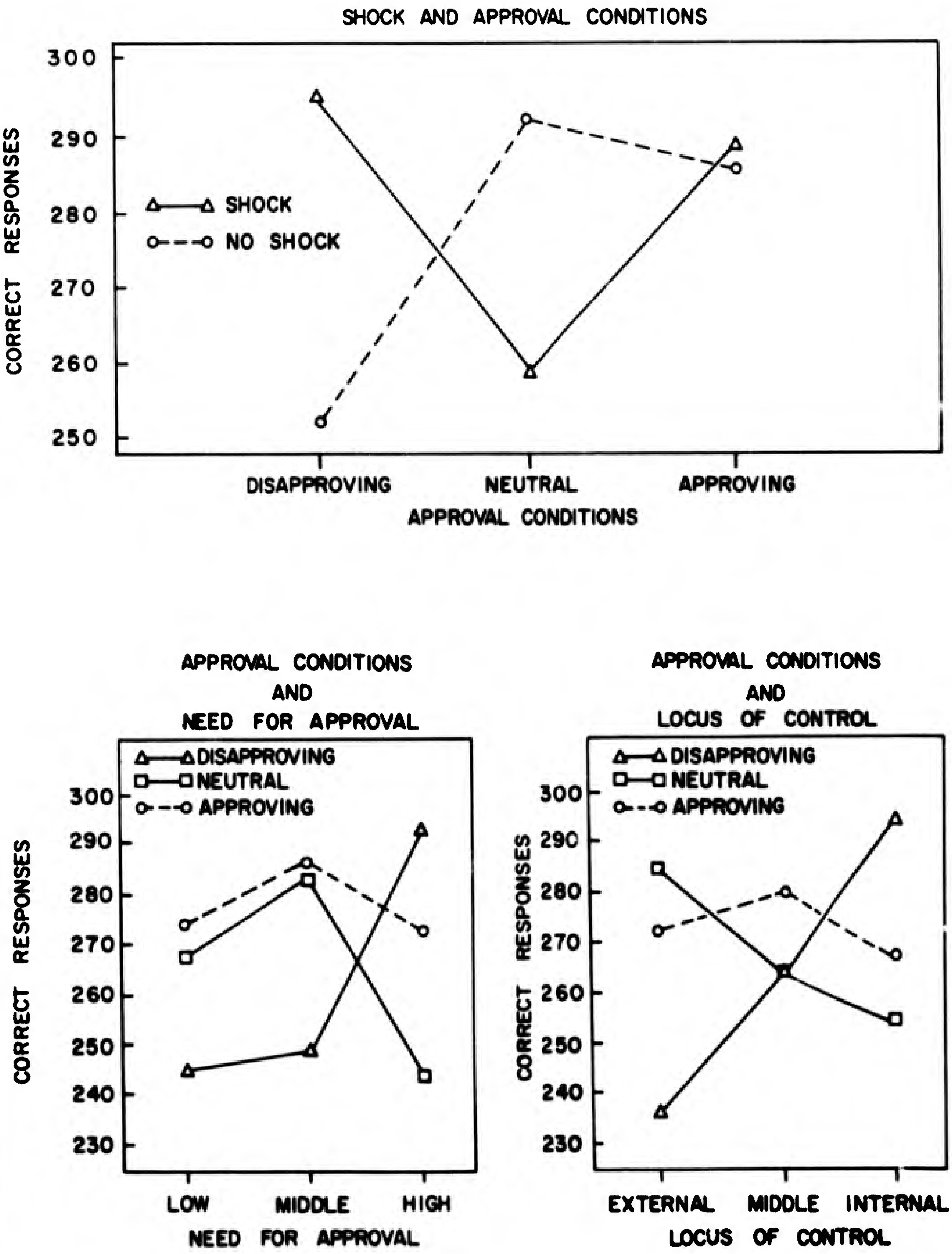
The effects of shock on subjects in the disapproving condition suggest that experimenter approval may be serving a cue as well as an incentive, function. Disapproval comments seemed to provide the subject with information about which of the two tasks he should try harder on. Prior to the disapproval comments, shocked subjects were presumably most concerned with the distractor task, while no-shock subjects were directing their efforts toward the more challenging verbal task. Disapproval presumably suggested to both groups that they were attending to the wrong task. Subjects in the disapproving condition sought to clarify why they were receiving criticism. Clear differences in performance appeared only after trials 8 and 12, the trials after which the disapproval comments were made.

Need for approval and locus of control effects were dependent on approval conditions. High M-C SD subjects seemed to be mobilized by the threat of disapproval, whereas disapproval had opposite effects on other subjects.

External-control subjects seemed to stop trying in the disapproving condition, while internal-control subjects seemed to try even harder. This would not necessarily be expected, from previous conceptions of locus of control. This phenomenon should be replicated before firm conclusions are drawn.

Figure 6

Mean Number of Correct Responses as a Function of Approval Conditions, Shock, and Test Scores, Study IX



STUDY X

Purpose

The primary purpose of this study was to collect criterion data for test development. In Study IX the different experimental conditions had vastly different effects on performance. Consequently, it was necessary to employ the same six conditions in collecting cross-validation data. This study also served, then, as a replication of Study IX.

To test the validity of the interpretation of the results of Study IX, the reaction-time scores of subjects were recorded, as well as performance on the verbal task. Reaction-time scores had not been recorded in Study IX, since the reaction-time task was employed simply as a way of rationalizing the use of shock as a Drive stimulus. Performance on the verbal task, however, suggested that shock served also as an incentive for the subject to attend to the reaction-time task, at the expense of verbal task performance. It also seemed that the experimenter's disapproving comments suggested to the subject that he was attending to the wrong task. If these phenomena were in fact responsible for the findings in Study IX, it would be expected that the best reaction-time performance would occur in the groups with the worst verbal task performance: the shock, neutral and no-shock, disapproving conditions. Reaction-time scores were recorded to test this.

Method

Apparatus

Only two changes in apparatus were made.

A clock was wired in series with the shock and buzzer circuits, to record reaction times. The clock was started by the same key with which the experimenter initiated the shock (or buzzer) stimulus, and was stopped when the subject pressed the response key.

The second apparatus change involved substituting a film strip projector, adapted to accommodate an endless loop, for the slide projector as a means of presenting the visual stimuli. The slide projector had presented some apparatus reliability problems which had necessitated discarding the results of some subjects. The film strip projector was substituted for this reason. A projection screen was employed with the film strip projector.

Subjects

The same subject pool, two freshman AFROTC sections, was employed for this study as for Study VIII and Study IV. Most subjects had already participated in these two studies, in that order, before being solicited for this study. Total number of subjects for this study was 108. Equipment failure necessitated discarding the results of 14 subjects; total useable number of subjects,then, was 94. All had been previously tested, in their respective AFROTC sections, on the MAS, M-C SD, LC, EPPS, ACL, EIS, and IS.

Experimental Design and Procedure

There were no changes in procedure or experimental design.

Results

When the subjects within each experimental condition were trichotomized on the bases of MAS scores, there were gross inequalities in numbers of subjects per category per condition. This was true also when M-C SD or LC scores were used to trichotomize the subjects, although the inequalities occurred in different treatment conditions in each case. Cell frequencies ranged from 1 to 10, despite the total of 94 subjects. Subjects had been randomly assigned to conditions without knowledge of test scores. The random assignment, by chance, produced these inequalities.

Analyses of variance (unweighted means) were conducted for each test-score categorization. The results of the three analyses were grossly discrepant in many places. Computations were rechecked and found correct. On inspection it was found that the several very low cell frequencies were responsible for the discrepancies in the analyses. Because of the unweighted means analysis of variance procedure, the small number of subjects in underrepresented cells contributed disproportionately to the overall variance. In each underrepresented cell there was, it was found, one subject with particularly deviant performance. These artifacts biased even the results which were not relevant to test-score categorizations. For example, the $\underline{F}$ value for the approval conditions main effect was 1.94 in the analysis employing M-C SD categories, but less than 1.0 in the LC analysis.

The statistical analyses were, therefore, grossly unreliable. For this reason they are not reported, nor was confidence placed in them in interpreting the results. Analyses could have been conducted disregarding test scores, or dichotomizing instead of trichotomizing subjects within treatment groups. Previous results, however, indicated that either of these procedures would also have produced ambiguous results. Individual differences in scores on these tests consistently interacted with the effects of treatment conditions, and middle-scoring subjects on each of these tests were often deviant, so as to contraindicate dividing subjects at the median.

Since this study was originally intended only as a replication, the results were simply inspected for reoccurrence of the primary results of Study IX. The most important finding was that overall performance scores were substantially lower in this study than in Study IX. Summing across trials and response dominance, mean performance was over 70 per cent in Study IX, but only about 60 per cent in this study.

Unsystematic observation of the subjects had suggested that all subjects this time believed that the reaction-time task was the important task, and the verbal learning task only a distractor task. This was, of course, what the subjects had been told in the instructions. The subjects seemed very concerned about their reaction-time performance, and very aware of the experimenter's resetting the clock and recording scores. The presence of the clock and the experimenter's attention to it in this study (the only difference in procedure from Study IX) were apparently unambiguous cues to all subjects that the reaction-time task was, as indicated in the instructions, the important task. This might explain the generally lower scores on the verbal task in this study, as well as some departures from the findings of Study IX.

Verbal task performance in this study was generally lower under shock than under no shock, in all three approval conditions. This trend was slight, and nonsignificant in all three analyses of variance. It does suggest, however, that the approving and disapproving comments in this study did not change the subject's assumption that the verbal task was, as stated, the less important task. Indirectly it also supports the assumption that the cue-function of approving and disapproving comments in Study IX was, as expected, responsible for many of the results. Shock depressed performance most in the neutral condition, and had least effect in the approving condition, as before; it also depressed performance in the disapproval condition, however, contrary to the finding in Study IX.

The findings regarding MAS, M-C SD, and LC scores were not replicated. The interactions of test scores with approval conditions were both nonsignificant and minimal, and were not of the nature found previously. There was, in fact, a very restricted range of individual differences in performance scores, compared to Study IX; and none of the test scores seemed related in any way to performance scores, in any condition.

Reaction-time scores were compared, group by group, with scores on the verbal task. Failure of reaction-time equipment made the scores unavailable for several subjects; and there was great intra-individual variability, trial by trial. The subject had to reach from chair - arm to table to press the reaction-time key, and sometimes the subject missed the key on his first pass.(This phenomenon was not apparent, at least for no-shock subjects, in Study IX.) There was an inverse relationship, group by group, between reaction-time scores and verbal task scores. Subjects in the shock, neutral condition had best reaction-time scores and worst verbal task scores; subjects in the no shock, neutral condition had worst reaction-time and best verbal task scores. Other groups were ordered between these two groups and the inverse relationship was consistent. Differences were slight, however, and these was much overlap in scores between groups.

Conclusions

There was no reliable basis for firm conclusions based on these results. The most obvious general conclusion was that even apparently minor changes of procedure can have major effects on performance, and major effects on the effects of other variables on performance. These results did nothing to reduce confidence in the interpretation of the results in Study IX. They can scarcely be said to have supported these interpretations either, however.

The results of this study were inspected, as those in Study IX, for replication of the findings of Chiles (ref. 39) or Ramond (ref. 40). Again, there was no clear evidence of Drive effects, early in performance, as mediated by response dominance. Drive phenomena again could easily have been obscured by the complex incentive and cue effects which seemed to have been predominant in affecting performance.

SUMMARY: MOTIVATION AND VERBAL LEARNING

Motivational Variables in Verbal Learning

Once more Drive effects on performance could not be detected, apparently because of supervening incentive effects. In the reaction-time and the finger-retraction conditioning tasks, the shock Drive stimulus served as an incentive which enhanced performance. In the verbal task the shock, as it was used, was an incentive for performance on a task other than the one of interest; it seemed, then, to detract from performance on the verbal task.

Approval incentives again seemed to be important, in Study IX. Approving comments counteracted the shock incentive to disregard the verbal task; and individual differences in need for approval were related to performance differences, although this relationship differed in different approval conditions. Individual differences in locus of control were also related to performance differences in Study IX, and again the relationship varied in different approval conditions.

Study X failed to replicate any of the findings in Study IX. Consequently, conclusions must be guarded concerning motivational variables in verbal learning. Under the circumstances, the results of Study X did not necessarily contradict the results of Study IX. The net effect of both studies was to illuminate the complexity of motivational phenomena in this task, and the great difficulty in truly replicating conditions of motivational importance.

Implications for Test Construction

Again, this task is not likely to be useful in constructing a test to measure individual differences in Drive. It may, as used in Study IX, be of use in devising tests of need for approval and locus of control.

However, as used in Study X it is not likely to be useful for constructing tests of these variables, either.

It is also obvious that item-analysis and cross-validation procedures need to be conducted separately within each of the six experimental conditions employed.

SECTION VII
MOTIVATIONAL CORRELATES OF CONCEPT FORMATION

RATIONALE AND METHOD

Rationale

There were three purposes for developing the concept-formation task. One was to provide a second verbal task, against which to compare results of the simple verbal paired-associates task. The second was to investigate motivational phenomena in a maximally complex task; in this case, the task represented complexity of mediating processes, rather than of perceptual or motor processes. The third purpose was to try again to devise a task which would be sensitive to individual differences in Drive.

In inspecting the results obtained with the simple verbal paired-associates task, a methodological problem became apparent. Of the two response words available to the subject, there was a clear difference in strength of association with the stimulus word. However, this difference may have been functionally obliviated after the first few triads were presented. The subject seemed to learn quickly that the obviously synonymous word was correct only half the time. The dominance of the tendency to emit that word could, then, have vanished rapidly. The subject learned, in effect, that the probability of correctness of each response word was 0.50.

Several subjects, in fact, gave more unrelated than synonymous words during several early trials. A few gave almost exclusively unrelated-word responses for several trials. This could scarcely indicate an originally stronger associative tendency. It seemed more likely that this represented a difference in "strategy." The subjects perceived it as more difficult, and a greater challenge, when the obviously associated response word was incorrect than when it was correct. Some subjects, then, apparently tried to learn the "hard" items first. Such strategy differences, if they can be assumed, are clearly of higher-order habit, rather than simple associative nature. They could be added to the list of higher-order habit variables which had apparently predominated over simple associative and Drive phenomena.

In developing the concept-formation task, an attempt was made to minimize the extent to which strategy variables would override simple associative tendencies.

Method

Stimulus Materials

The concept-formation task was also a verbal associative-learning task, with the stimuli visually presented.

The stimuli and responses were the names of six colors: blue, black, brown, red, green, and yellow. For each stimulus, the subject had six, rather than two, responses to choose from. However, the correct response was not the same color-name as that represented by the stimulus.

In addition, each of the stimulus color-names was printed in each of the six different colors - blue, black, etc. The color in which the stimulus was printed was also not the correct response. For example, the response word "brown" was correct for the stimulus word "yellow" printed in green. It was assumed that there would be strong previously learned associative tendencies to either give the color-name represented by the stimulus (ie, to read the stimulus word) or to name the color in which the stimulus was printed. Since the response words were drawn from the same population of verbal elements, competition among response alternatives was expected to be maximal. The chance probability that any of the six responses would be correct was 0.167; it was assumed that this would maximize the interference from the strong previously learned incorrect response tendencies.

Two other restrictions were made in selecting the correct answer for each stimulus word. "Red" was never the correct answer for "green," and vice versa, since these colors are seen as a pair more often than the other possible color combinations. Also, if "green" were the correct answer for the word "brown," then the reverse was not true, ie, "brown" was not the correct answer for the word "green."

Finally, the subject was not told whether the correct response depended on (was the predetermined associate of) the word represented by the stimulus or the color in which the stimulus was presented. The subject had to learn first which aspect of the stimulus was the appropriate cue, and then learn the correct associative responses.

Pilot testing confirmed that this was a very difficult task. In the original version of the task there were eight, rather than six, color-names. This proved to make the task too difficult. Very few subjects could learn the task within 50 minutes, and many never attained better than a chance level of performance.

Apparatus

A slide projector and projection screen were used to project the stimuli. Color fidelity of the slides was sufficient to permit easy recognition of the color in which the stimulus word was presented. Each stimulus was presented for a duration of 7.83 sec. Interstimulus interval was 0.7 sec. Both intervals were controlled by the automatic timer on the slide projector.

Procedure

Subjects were told that this was a complex concept-formation task. (Complete instructions and details of apparatus, procedure, and results are available in ref. 42.) To explain the nature of the task, two practice tasks were administered. In these, responses were the words "circle," "square," and "triangle." These words were also the stimuli. They were written on a blackboard for the subject; and each word was enclosed by a geometric form (circle, square, or triangle) other than that denoted by the stimulus word. The correct response in one practice task depended on the stimulus word, in the other on the geometric form enclosing the word. In each case, the correct response was by chance the same as the irrelevant aspect of the stimulus one time in three, but never the same as the relevant aspect of the stimulus. (In the experimental task the chance coincidence was one in six.)

Practice was continued until the subject had both learned the correct responses and could verbalize the principle for learning these responses (ie, that one aspect of the stimulus was irrelevant and that the correct response was not the word which was represented by either aspect of the stimulus except by coincidence). The subject then started the experimental task, without being told which aspect of the stimulus would be relevant and which irrelevant.

The 36 stimulus slides were arranged in a random order with the following restrictions. Each consecutive group of six slides contained each of the six stimulus words once and only once. Each group also contained each of the six stimulus colors once and only once. The entire series of 36 slides used each of the 36 possible combinations of stimulus word and stimulus color once and only once. There were six slides in which the stimulus word was the same as the color in which that word was printed. These were randomly distributed in the series of 36 slides. The series of 36 slides was presented four times, in the same order. Each successive series was considered a trial.

STUDY XI

Purpose

The purpose of this study was to continue investigating the utility of this task for exploring individual differences in Drive level.

Method

Subjects

The same subject pool was used as in studies IV, VIII, and X. Subjects had taken the MAS, M-C SD, LC, EPPS, ACL, EIS, and IS in their freshman AFROTC classes. A total of 56 subjects were used.

Apparatus

The apparatus described above was employed.

Experimental Design

A control task was used to determine the effect of the irrelevant cue. The same stimulus words were used, but all were printed in black. As in the experimental task, the stimuli were presented in random order, with each stimulus appearing once in each set of six.

One third of the subjects took only the control task. The rest of the subjects took the experimental task twice. One time the stimulus word was the relevant cue (word task) and the other time the color the stimulus word was printed in was the relevant cue (color task). One group of subjects took the word task first, followed by the color task; the other group took the tasks in reverse order.

Results with the control task were analyzed with results with the word task. The task order factor was control task only, word task first, and word task second. In analysis of the color task, treatment order groups were color task first and color task second.

Order of practice tasks was counterbalanced. Half the subjects took the word practice task first, followed by the form practice task (in which the relevant aspect of the stimulus was the geometric form enclosing the stimulus word). The rest of the subjects took the practice tasks in reverse order. This defined the practice task order factor.

Within each of the combinations of task order and practice task order, subjects were dichotomized as high- or low-anxious on the basis of MAS scores. This defined the anxiety factor. The experimental design for analysis of performance on the word task was, then, a 3 x 2 x 2 x 4 (task order x practice task order x anxiety x trials) factorial analysis of variance (unweighted means), with repeated measures over trials. For the color task the design was a 2 x 2 x 2 x 4, like the above except for omission of the control task from the task order factor.

After these analyses using MAS scores, subjects were repooled within treatment conditions and redichotomized on the basis of M-C SD scores. The analyses were repeated, substituting the M-C SD test score factor for the anxiety factor. This was repeated using LC scores to dichotomize subjects.

Because of the obvious intellectual demand of this task, the relation of performance to intelligence was also examined. Scores on the SAT, administered in freshman orientation, were obtained for those subjects for whom they were available. (Eight subjects in this study had not taken the SAT, of whom two were in the control-task-only group.) Analyses were repeated using the SAT verbal (SAT-V) to categorize subjects, and then using SAT mathematical (SAT-M) scores.

The dependent variable in all analyses was number of errors.

Results

Summaries of the analyses of variance are in table XI.

There was a distinct "learning to learn" phenomenon. On both the word and the color task, there were more errors if the task was the first one taken than if it was the second (task order, $p < .01$). Learning was generally gradual on the first task taken but rapid on the second (task order x trials, $p < .01$ in most analyses). On the second task taken, individual differences were minimal. Consequently, almost all phenomena of interest were apparent in interactions with task order. Even when there were differences on the first task taken these differences were minimal on the second task.

There were extreme individual differences, however, on the first task taken, whether it was the word task or the color task. Of the 40 subjects who took the two experimental tasks, 11 never reached errorless performance level on the first task taken, and a few scarcely improved beyond chance level through all 4 trials. At the other extreme, 2 subjects reached errorless performance within 1 trial, and several others reached this level within 2 trials.

Some of the results, although statistically significant, appeared to be an artifact of these extreme individual differences. Since there were few subjects per treatment condition, the presence of one or two deviant subjects had a major effect on the group mean. This was manifest, for example, in interactions involving practice-task order, all of which seemed to reflect only artifacts of the presence of one or two deviant subjects in a particular group.

The results of importance were those involving test score x task order x trials interactions. In all cases, differences between groups were clear only on the middle trials of the task taken first. Differences were minimal on the task taken second. They were also minimal during the first trial (when almost all subjects were still performing at near chance level), and the last trial (when almost all subjects had few if any errors), of the task taken first.

High-anxious subjects made more errors (MAS test score x task order x trials, $p < .01$) than low-anxious subjects on the word task, during middle trials when the word task was taken first. As noted, differences were slight on early and late trials, and slight when the word task was taken second. On the color task there was the same trend, but it was non-significant.

TABLE XI

SUMMARIES OF ANALYSES OF VARIANCE OF ERRORS, STUDY XI

Word Task: Source of Variation	df	M A S MS	M A S F	M-C SD MS	M-C SD F	L C MS	L C F	df	SAT - V MS	SAT - V F	SAT - M MS	SAT - M F
Between Subjects	55							47				
Test Score (TS)	1	2.9	1	0.4	1	0.4	1	1	15.4	2.79	9.2	1.6
Practice Task Order (PTO)	1	1.4	1	3.2	1	4.4	1	1	4.2	1	3.7	1
Task Order (TO)	2	4 4	10.8**	59.1	12.9**	70.7	14.2**	2	60.1	10.8**	56.9	9.9**
TS x PTO	1	3.5	1	8.8	1.9	0.0	1	1	0.0	1	0.6	1
TS x TO	2	8.8	2.0	6.2	1.4	2.2	1	2	3.3	1	2.5	1
PTO x TO	2	1.9	1	2.0	1	2.3	1	2	2.2	1	3.2	1
TS x PTO x TO	2	0.6	1	0.5	1	0.5	1	2	0.8	1	2.6	1
(Subj. within grp. Error)	(44)	(4.4)		(4.6)		(5.0)		(36)	(5.5)		(5.7)	
Within Subjects	1288							1104				
Trials (Tr)	23	21.0	89.6**	20.4	80.8**	20.3	77.2**	23	20.2	65.9**	21.0	72.4**
TS x Tr	23	0.2	1	0.2	1	0.2	1	23	0.3	1	0.7	2.3**
PTO x Tr	23	0.2	1.0	0.2	1	0.2	1	23	0.2	1	0.2	1
TO x Tr	46	0.8	3.5**	1.0	3.8**	1.0	3.8**	46	0.9	2.9**	1.0	3.4**
TS x PTO x Tr	23	0.2	1	0.6	2.4**	0.1	1	23	0.1	1	0.2	1
TS x TO x Tr	46	0.6	2.7**	0.2	1	0.2	1	46	0.3	1	0.3	1.2
PTO x TO x Tr	46	0.3	1.1	0.3	1.1	0.2	1	46	0.3	1	0.3	1
TS x PTO x TO x Tr	46	0.1	1	0.2	1	0.2	1	46	0.1	1	0.2	1
(Tr x Subj. within grp. Error)	(1012)	(0.2)		(0.2)		(0.3)		(828)	(0.3)		(0.2)	

* p<.05

** p<.01

TABLE XI (CONTINUED

Color Task: Source of Variation	df	MAS MS	MAS F	M-C SD MS	M-C SD F	LC MS	LC F	df	SAT-V MS	SAT-V F	SAT-M MS	SAT-M F
Between Subjects	39							33				
Test Score (TS)	1	0.6	1	4.1	1.0	2.3	1	1	11.3	2.5	13.1	2.2
Practice Task Order (PTO)	1	8.8	1.9	1.6	1	12.6	3.3	1	3.6	1	6.9	1.2
Task Order (TO)	1	88.5	18.7**	110.2	28.4**	77.8	20.3**	1	71.1	15.6**	107.2	18.1**
TS x PTO	1	16.3	3.5	2.7	1	37.5	9.8**	1	0.8	1	7.2	1.2
TS x TO	1	1.5	1	3.3	1	11.1	2.9	1	32.2	7.1*	4.5	1
PTO x TO	1	0.3	1	2.4	1	0.0	1	1	1.3	1	0.0	1
TS x PTO x TO	1	0.1	1	20.0	5.1*	1.0	1	1	4.5	1	4.1	1
(Subj. within grp. Error)	(32)	(4.7)		(3.9)		(3.8)		(26)	(4.6)		(6.0)	
Within Subjects	920							782				
Trials (Tr)	23	19.5	71.8**	18.2	76.2**	19.5	75.5**	23	18.9	66.2**	16.9	50.7**
TS x Tr	23	0.3	1.0	0.2	1	0.3	1.2	23	0.7	2.4**	0.3	1
PTO x Tr	23	0.3	1.0	0.3	1.5	0.2	1	23	0.3	1	0.3	1
TO x Tr	23	0.4	1.5	0.5	1.9**	0.3	1.1	23	0.3	1	0.4	1.2
TS x PTO x Tr	23	0.2	1	0.7	2.7**	0.4	1.6*	23	0.1	1	0.3	1
TS x TO x Tr	23	0.3	1.2	0.1	1	0.5	2.0**	23	0.4	1.4	0.4	1.2
PTO x TO x Tr	23	0.2	1	0.1	1	0.2	1	23	0.2	1	0.2	1
TS x PTO x TO x Tr	23	0.4	1.3	0.4	1.6	0.7	2.9**	23	0.3	1	0.2	1
(Tr x Subj. within grp. Error)	(736)	(0.3)		(0.2)		(0.3)		(598)	(0.3)		(0.3)	

* p<.05
** p<.01

Similarly, external-control subjects made more errors (LC test score x task order x trials, p<.01) than internal-control subjects on the color task, during middle trials when the color task was taken first. Again, differences were slight on early and late trials, and when the color task was taken second. Differences were also slight on the word task.

Need for approval scores were not related to error frequency except in interactions involving practice-task order. As mentioned, these interactions appeared to be artifacts of the presence of one or two deviant subjects, when classified by M-C SD scores, in groups which had different orders of practice and test tasks. Practice-task order, as expected, had little general effect.

SAT verbal scores were inversely related to errors on the color task, and SAT mathematical scores on the word task. Again, these differences were reflected in significant interactions with task order and/or trials. This raised question whether the relationship to error frequency of MAS and LC scores was attributable to differences in intelligence correlated with MAS and LC scores. SAT verbal and mathematical scores were, consequently, intercorrelated with scores on the MAS, M-C SD, and LC (see table XII). Neither verbal nor mathematical scores were correlated with MAS, M-C SD, or LC scores. Anxiety and locus of control seemed, then, to be independent of intelligence in affecting performance.

Finally, performance on the control task was better than performance on the experimental task, even when the experimental task was taken second. This is reflected in the task order main effect (p<.01) in analyses of performance on the word task. The irrelevant aspect of the stimulus did detract from performance.

TABLE XII
INTERCORRELATIONS OF SAT SCORES WITH MAS, M-C SD, AND LC SCORES
STUDY XI

	M-C SD	LC	SAT-V	SAT-M
MAS	-.51**	-.33*	-.12	-.11
M-C SD		.05	-.08	-.13
LC			.16	.14
SAT-V				.62**

* p<.10
** p<.01

Conclusions

These results suggested that this task might in fact be sensitive to Drive effects. It seemed safe to assume that the correct response tendency was weaker than one or more incorrect tendencies. High-anxious subjects made more errors than low-anxious subjects, as would be anticipated from Drive theory.

The difference between anxiety groups was manifest only during middle trials on the first task taken. However, in retrospect, this was the only point where clear differences should have been expected. Improvement in performance was too uniformly rapid on the second task taken to permit clear individual differences. On the first task taken, performance during most of the first trial stayed near chance level for almost all subjects. At this point the subjects were attempting to learn which aspect of the stimulus was relevant and then seeking, randomly, to discover the correct associative responses. During the last trial, many subjects reached errorless performance, and individual differences in errors were again reduced.

The difference between anxiety groups was nonsignificant on the color task. Again, however, this was not surprising. Only one third of the subjects took this task first. Since the number of subjects was small, and subjects were simply dichotomized at the median on MAS scores, the nonsignificance could be expected on the basis of lack of statistical power. The trend in anxiety differences was the same on the color task as on the word task, and on the latter the difference was statistically significant.

The difference attributable to LC scores was also of interest. This was again buried in interactions, like the anxiety effect, and again significant on only one task. It still appeared to be, quite possibly, a reliable phenomenon. Several subjects seemed to "give up" after one or two trials of trying unsuccessfully to learn the associative responses. From that time on they seemed to respond more or less randomly. This would be consistent with expectations from Social Learning Theoryof the performance of external-control subjects.

Also of interest was the lack of relationship of M-C SD scores to performance. Need for approval differences had been related, with striking consistency, to performance on other tasks. However, at the risk of overinterpreting negative results, this, too, was not surprising. This task elicited great interest and effort from all subjects. It was an obvious intellectual challenge. Achievement incentives seemed to be quite adequate to elicit maximum effort from all subjects. Approval incentives could, then, have been superfluous.

It was also unsurprising to find differences in intelligence related to performance. Since intelligence scores were not correlated with MAS or LC scores, this relationship did not compromise the findings regarding anxiety and locus of control.

Finally, it was obvious that only one task, the word or the color task, need be given to each subject. Differences on the second task could be expected to be minimal. In addition, it was desirable to hold practice-task order constant, since this had no overall effect.

STUDY XII

Purpose

Study XI suggested that this task might be the only one, of all those used in this research, which was sensitive to individual differences in chronic Drive level. The purpose of this study was to further test the extent to which Drive increase inhibits performance on this task.

The specific purpose was to determine the effects of experimentally manipulated Drive. MAS scores were again collected, but only as a secondary interest. It had become quite apparent that the MAS is contaminated by differences in non-Drive variables. A more powerful test of Drive effects would be provided by a different operation (namely, an experimental operation) for defining Drive.

Locus of control effects were not investigated. These were of interest, and also in need of replication. Nevertheless, it seemed far more important to pursue the possibility that this task could, in future research, be used to construct a test sensitive to individual differences in Drive.

It was not possible, in the time remaining for the present research project, to collect extensive criterion data for item-analyses or cross-validation.

Method

Apparatus

The same apparatus for projecting stimuli was used as in Study XI.

The apparatus used to manipulate Drive was the shock apparatus used in studies VI, VII, and VIII, the finger-retraction conditioning studies. Both electrodes were taped to the subject's fingers (the classical conditioning procedure) so that the shock was unavoidable.

Subjects

Subjects were 36 volunteers from an introductory psychology class, randomly assigned to treatment conditions. Subjects completed the MAS before taking the concept-formation task.

Experimental Design

Subjects took only the color task. As before, testing was continued for four trials (144 stimulus presentations). The experimental design was a 3 x 4 (conditions x trials) analysis of variance, with repeated measures over trials. Dependent variable was number of errors.

Procedure

Three treatment conditions were employed. In the control condition, the procedure was identical to that of Study XI; no use of or reference to the shock apparatus was made. In the shock condition, the electrodes were taped to the subject's fingers, and the subject was told simply that the experimenter was interested in the effect of stress on performance. The shock-incentive condition was identical to the shock condition except that the subject was told that he would receive a mild electric shock whenever it seemed he was not "doing as well as we think you should." This condition was included to control for incentive effects of shock. In the shock condition the shock might serve as an incentive for disregarding the task, whereas in the shock-incentive condition it was expected to be an incentive for maximum effort on the task.

Shock was delivered in the interstimulus interval, in an effort to minimize the extent to which it would have a disrupting effect on performance.

Results

The summary of the analysis of variance is in table XIII, as is the table of mean number of errors per subject in each condition.

Both shock conditions significantly increased errors (conditions main effect, $p < .05$). There was only a minimal difference between the two shock conditions. In both shock conditions, performance reached only about the 50 per cent correct level on the last trial; in the control condition there were only approximately 30 per cent errors, overall, on the last trial.

Individual differences within conditions were again substantial. In the control conditions, 5 of the 12 subjects had no errors on the last trial; and 3 reached errorless performance toward the end of the last trial; the remaining 4 subjects were performing near chance level even at the end of all 4 trials. In the shock-incentive condition, one subject was errorless in the last trial, and one more reached that level toward the end of the last trial; 6 were still near chance level, and the other 4 were still not close to errorless performance. In the shock condition, 3 subjects reached errorless performance toward the end of the last trial, 7 were still near chance, and 2 were intermediate.

Discussion

Shock markedly interfered with performance, as expected, and attributing an incentive function to the shock did not reduce its effect. This was not wholly unexpected, despite the fact that the incentive function of shock had so consistently affected performance on other tasks. The shock used in this study was mild, and all subjects, in all conditions, seemed to try to do the best they could on the task. Additional incentives could, then, have been superfluous.

These results further supported further use of this task for investigating the effects of Drive, including individual differences in Drive, on performance.

It is possible that the effect of shock was attributable to its distractor, or disrupting, influence, rather than to Drive. This could not be ascertained from this study. Further research is needed to separate disrupting and Drive effects. Nevertheless, the results of this study were consistent with MAS findings in Study XI. MAS scores were inspected in this study, and showed a trend similar to that of Study XI. There were too few subjects to include anxiety as a within-treatment factor, however; and the relationship of MAS scores to errors was not perfectly consistent.

TABLE XIII

SUMMARY OF ANALYSIS OF VARIANCE AND TABLE OF MEANS OF ERROR SCORES, STUDY XII

Source of Variation	df	Mean Square	F
Between Subjects	35		
Treatment Conditions (c)	2	797.7	4.52*
(Subj. within grp. Error)	(33)	(176.3)	
Within Subjects	108		
Trials (Tr)	3	1306.0	44.33**
C x Tr	6	48.2	1.64
(Tr x Subj. within grp. Error)	(99)	(29.5)	

* $p<.05$
** $p<.01$

Mean Number of Errors per Subject for Each Group

Condition	Mean Errors per Subject
Control	74.75
Shock	105.25
Shock-Incentive	100.00
(Chance Level of Performance)	(120.00)

SUMMARY: MOTIVATION AND CONCEPT-FORMATION

Motivational Variables in Concept-Formation

Considerably more research is necessary before it can be confidently concluded that this task is sensitive to individual differences in Drive. Results with the MAS should be replicated, employing greater numbers of subjects and categorizing subjects into more than just a high- and a low-anxious group. Non-Drive variables related to MAS scores (eg, social desirability response-set, generalized response inhibition or nonfunctional responding under stress, etc) will, even then, leave some ambiguity in results with the MAS.

Similarly, other experimental operations for manipulating Drive should be employed. These should include operations which are not likely to have distracting effects, through their stimulus properties, or represent incentives to attend less to the task. When Drive is manipulated through application of noxious stimuli in the task situation, however, it may be difficult to avoid distracting or incentive contamination of Drive effects. Ironically, experimental operations may be even more contaminated than test measures in this instance.

Nevertheless, the results available indicate that this may be closer than any task so far devised to a task which is sensitive to the effects of individual differences in Drive.

It is regrettable that this task was not devised early enough in the research to permit extensive collection of data with it to devise a (possibly) more powerful measure of Drive differences. This effort would seem to be well worthwhile in future research.

Locus of control differences also seemed, in Study XI, to affect performance on this task. Again, this finding should be replicated, and should be further tested by experimental manipulation of locus of control. The effect observed was, however, consistent with expectation.

Incentive variables seemed ineffective in this task. Need for approval was unrelated to performance in Study XI, and the shock incentive was ineffective in Study XII. This seemed, however, simply to reflect the potent built-in achievement incentive in the task, rather than the irrelevance of incentives. The effects of incentives apparently do asymptote at some level, beyond which further incentives do not enhance performance.

Implications for Test Construction

Data from Study XI should be useful for constructing tests of Drive and locus of control differences. The most useful data for this purpose were those relating to performance on the first task the subject took, whether word or color task. Data from the control task appeared to be useless in test development. In Study XII there were too few subjects per condition to render very useful data for cross-validations, and the effects of the shock conditions might well have obscured the effects of individual differences in motivational variables.

SECTION VIII
TEST CONSTRUCTION AND CROSS-VALIDATION

PROCEDURE FOR TEST CONSTRUCTION

A separate test battery was developed for each of the four kinds of tasks used in the research (reaction time, finger-retraction conditioning, simple verbal associative learning, and concept formation). It did not seem likely that an onmibus battery would prove to be feasible. The procedure in developing each of these batteries was as follows.

One group of subjects was selected to provide criterion data for item analysis. In selecting this group, the aim was to maximize the probability that the items selected would be optimally useful in research employing that type of task. Consequently, the group selected was that which was tested under most nearly modal conditions.

For the reaction-time test battery, latency on the simple task was chosen as the criterion, and subjects were those in Study IV who were tested on the simple task first (Order 1), without shock. For the finger-retraction task, subjects in the low shock, classical procedure in Study VI were used. Since results were extremely similar for the latency and frequency measures in this study, both measures were used as criteria. An item was retained if it discriminated on either criterion, unless it discriminated with opposite keying on the other criterion. (Very few items had to be discarded because of discriminating in opposite directions for the two criteria.) For the simple paired-associates battery, the criterion employed was performance on all items, across all trials. No distinction was made between performance on dominant-response-correct and dominant-response-incorrect items, since original response dominance did not seem to be differentially related to motivational variables. Subjects used were those in Study X who were tested under no shock, neutral approval conditions. For the concept-formation task battery, total number of errors on the first task taken was the criterion. All subjects used in Study XI, except for those who took only the control task, were pooled for the item analysis.

The item analyses were then performed. Because of the small numbers of subjects in each group, a sequential analysis (ref. 33) item analysis procedure was employed. An alpha level of 0.12 was used as the criterion for item-selection. Items were also retained if they were in the "no decision" category (neither accepted nor rejected).

RESULTS OF ITEM ANALYSES

The items selected were compiled into short-form tests. Item keying was determined by results of the item analyses, disregarding the original keying of the item on the test from which it was drawn. This resulted in a short form of each of the tests used with these subjects: the MAS, M-C SD, etc.

The original test items used are reproduced in table XIV, with items keyed as in the original tests. In table XV are the items retained in the short forms of these tests, keyed as indicated by item analyses. Where the short-form keying differed from the original, this is indicated by an asterisk. It should be noted that items were sometimes keyed differently in different short-form batteries.

PROCEDURE FOR CROSS-VALIDATIONS

The short forms were then used to rescore the test protocols of all subjects. The subject was given a score for the short form of each test, and also a score for the sum of the scores on the separate short forms. The summed short-form scores were denoted "battery" scores.

Correlations of these test scores with performance criterion scores were then computed, by Pearson $\underline{r}$ (see table XVI). Each test score and the total battery score were correlated with each performance criterion (eg, frequency and latency). In addition, each criterion score was correlated with the total score on the original (long form) tests, eg, the MAS and M-C SD. This was done to indicate the extent to which the short form decreased or increased the predictive power of the original test. These correlations were computed separately for the subjects in each experimental condition, in each study.

Correlations for the group which was used for item analysis are parenthesized; these correlations are, of course, spuriously high.

No correlations are reported for subjects in Study I, since no paper and pencil tests were used in this study, nor for Study VII, because of the special purpose and low number of subjects of that study. No correlations were computed with number of errors on the simple reaction-time task, because of the infrequency or errors on that task.

TABLE XIV
ITEMS FROM ORIGINAL TESTS, KEYED AS IN ORIGINAL

MAS

1. I do not tire quickly. (F)
2. I am troubled by attacks of nausea. (T)
3. I believe I am no more nervous than most others. (F)
4. I have very few headaches. (F)
5. I work under a great deal of tension. (T)
6. I cannot keep my mind on one thing. (T)
7. I worry over money and business. (T)
8. I frequently notice my hand shake when I try to do something. (T)
9. I blush no more often than others. (F)
10. I have diarrhea once a month or more. (T)
11. I worry quite a bit over possible misfortunes. (T)
12. I practically never blush. (F)
13. I am often afraid that I am going to blush. (T)
14. I have nightmares every few nights. (T)
15. My hands and feet are usually warm enough. (F)
16. I sweat very easily even on cool days. (T)
17. Sometimes when embarrassed, I break out in a sweat which annoys me greatly. (T)
18. I hardly ever notice my heart pounding and I am seldom short of breath. (F)
19. I feel hungry almost all the time. (T)
20. I am very seldom troubled by constipation. (F)
21. I have a great deal of stomach trouble. (T)
22. I have had periods in which I lost sleep over worry. (T)
23. My sleep is fitful and disturbed. (T)
24. I dream frequently about things that are best kept to myself. (T)
25. I am easily embarrassed. (T)
26. I am more sensitive than most other people. (T)
27. I frequently find myself worrying about something. (T)
28. I wish I could be as happy as others seem to be. (T)
29. I am usually calm and not easily upset. (F)
30. I cry easily. (T)
31. I feel anxiety about something or someone almost all of the time. (T)
32. I am happy most of the time. (F)
33. It makes me nervous to have to wait. (T)
34. I have periods of such great restlessness that I cannot sit long in a chair. (T)
35. Sometimes I become so excited that I find it hard to get to sleep. (T)
36. I have sometimes felt that difficulties were piling up so high that I could not overcome them. (T)
37. I must admit that I have at times been worried beyond reason over something that really did not matter. (T)
38. I have very few fears compared to my friends. (F)
39. I have been afraid of things or people that I knew could not hurt me. (T)
40. I certainly feel useless at times. (T)

41. I find it hard to keep my mind on a task or job. (T)
42. I am usually self-conscious. (T)
43. I am inclined to take things hard. (T)
44. I am a high-strung person. (T)
45. Life is a strain for me much of the time. (T)
46. At times I think I am no good at all. (T)
47. I am certainly lacking in self-confidence. (T)
48. I sometimes feel that I am about to go to pieces. (T)
49. I shrink from facing a crisis of difficulty. (T)
50. I am entirely self-confident. (F)

M-C SD

1. Before voting I thoroughly investigate the qualifications of all the candidates. (T)
2. I never hesitate to go out of my way to help someone in trouble. (T)
3. It is sometimes hard for me to go on with my work if I am not encouraged. (F)
4. I have never intensely disliked anyone. (T)
5. On occasion I have had doubts about my ability to succeed in life. (F)
6. I sometimes feel resentful when I don't get my way. (F)
7. I am always careful about my manner of dress. (T)
8. My table manners at home are as good as when I eat out in a restaurant. (T)
9. If I could get into a movie without paying and be sure I was not seen I would probably do it. (F)
10. On a few occasions, I have given up doing something because I though too little of my ability. (F)
11. I like to gossip at times. (F)
12. There have been times when I felt like rebelling against people in auth even though I knew they were right. (F)
13. No matter who I'm talking to, I'm always a good listener. (T)
14. I can remember "playing sick" to get out of something. (F)
15. There have been occasions when I took advantage of someone. (F)
16. I'm always willing to admit it when I make a mistake. (T)
17. I always try to practice what I preach. (T)
18. I don't find it particularly difficult to get along with loud mouthed, obnoxious people. (T)
19. I sometimes try to get even rather than forgive and forget. (F)
20. When I don't know something I don't at all mind admitting it. (T)
21. I am always courteous, even to people who are disagreeable. (T)
22. At times I have really insisted on having things my own way. (F)
23. There have been occasions when I felt like smashing things. (F)
24. I would never think of letting someone else be punished for my wrongdoings. (T)
25. I never resent being asked to return a favor. (T)
26. I have never been irked when people expressed ideas very different from my own. (T)

27. I never make a long trip without checking the safety of my car. (T)
28. There have been times when I was quite jealous of the good fortune of others. (F)
29. I have almost never felt the urge to tell someone off. (T)
30. I am sometimes irritated by people who ask favors of me. (F)
31. I have never felt that I was punished without cause. (T)
32. I sometimes think when people have a misfortune they only got what they deserved. (F)
33. I have never deliberately said something that hurt someone's feelings. (T)

LC

I more strongly believe that:

1. a) many times the reactions of teachers seem haphazard to me.
   b) in my experience I have noticed that there is a usually direct connection between how hard I study and the grades that I get.
2. a) I would be surprised if I discovered that social success is mostly a matter of chance.
   b) in our society social recognition has little to do with ability.
3. a) in my case making friends depends on how hard I work at it, luck has little or nothing to do with it.
   b) making friends is a funny business, it is so dependent on the right combination of circumstances.
4. a) through discussion I can change other people's opinions.
   b) whether or not a person will do what I want, depends mostly on how he happens to feel at the time.
5. a) I feel increasingly helpless in the face of what is happening in the world today.
   b) I sometimes feel personally to blame for the sad state of affairs in our government.
6. a) when I make plans, I am almost certain that I can make them work.
   b) it is not wise to plan too far ahead because most things turn out to be a matter of good or bad fortune anyhow.
7. a) in most cases the student, not the teacher, determines the grade.
   b) it seems many times that the grades one gets in school are more dependent on the instructor's whims than on what a student can really do.
8. a) making a lot of money is largely a matter of getting the right breaks.
   b) promotions are earned through hard work and persistence.
9. a) marriage is largely a gamble.
   b) the number of divorces indicates that more and more people are not trying to make their marriages work.
10. a) it is silly to think that one can really change another person's basic attitudes.
    b) when I am right I can convince others.
11. a) with enough effort we can wipe out political corruption.
    b) it is difficult for people to have much control over the things politicians do in office.

12. a) I have usually found that what is going to happen will happen, regardless of my actions.
    b) trusting to fate has never turned out as well for me as making a decision to make a definite course of action.
13. a) if one gets the right teacher he can do well, otherwise it is hopeless.
    b) the marks I get in class are completely my own responsibility.
14. a) as far as I am concerned becoming a success in our society is a matter of struggle, luck has little or nothing to do with it.
    b) getting a good job largely depends upon being in the right place at the right time.
15. a) people are longely because they don't know how to be friendly.
    b) making friends is largely a matter of being lucky enough to meet the right people.
16. a) changing people's opinions is often a hard job, but with enough effort it can be done.
    b) in general other people will do as they please in spite of my efforts to get them to listen to me.
17. a) as far as international affairs are concerned, most of us are the victims of forces we cannot understand, let alone control.
    b) by active participation in political and social affairs the people can control world events.
18. a) most people don't realize the extent to which their lives are controlled by accidental happenings.
    b) there really is no such thing as "luck".
19. a) if one wants to badly enough, he can overcome almost any obstacle in the path of academic success.
    b) some teachers will give you a poor grade no matter how hard you work.
20. a) getting promoted is really a matter of being a little luckier than the next guy.
    b) in our society a man's future earning power is dependent upon his ability.
21. a) anyone can have good friends if he knows how to interact with people.
    b) being able to get along with people seems to be more a matter of the other person's moods and feelings at the time rather than one's own actions.
22. a) I have little influence over the way other people behave.
    b) if one knows how to deal with people they are really quite easily led.
23. a) changing social attitudes is a tremendous undertaking, but every little bit helps.
    b) people being the way they are, some form of racial prejudice is inevitable.
24. a) people are responsible for their actions, both good and bad.
    b) many people could be described as victims of circumstances beyond their control.
25. a) sometimes I feel that I have little to do with the grades I get.
    b) in my case the grades I make are the result of my own efforts, luck has little or nothing to do with it.

55. a) for the most part teachers give students what they have earned.
b) taking an objective test is a lot like gambling, if you're lucky you make the right choices.

56. a) in the long run the socially undesirable or inadequate individuals reach their proper level in our society.
b) there's little use in trying very hard since keeping one's job seems more dependent on economic conditions than on one's abilities.

57. a) it is up to the person who isn't liked to figure out why.
b) people are so unpredictable, that it is hard to really get to know them.

58. a) it would surprise me to learn that a good many people in positions of authority are there largely as a matter of luck and not because of any special talents they have.
b) luck is an essential ingredient for rising to a position of authority.

59. a) somehow all the effort people put in trying to change social prejudices, doesn't seem to get anywhere.
b) in the long run people control the politicians, not vice versa.

60. a) in the long run the bad things that happen to us are balanced by the good ones.
b) most misfortunes are the result of lack of ability, or ignorance, or laziness, or all three.

IS

1. I like to take a chance just for the excitement. (T)
2. In the morning I usually bound out of bed energetically. (T)
3. I spend much of my leisure time out of doors. (T)
4. I usually think before I leap. (F)
5. I like mathematics. (F)
6. I answer questions quickly. (T)
7. I like to work crossword puzzles. (F)
8. I change my plans often. (T)
9. I like detailed work. (F)
10. I make up my mind quickly. (T)
11. As a youngster I enjoyed taking part in reckless stunts. (T)
12. I frequently forget things. (T)
13. I like to solve complex problems. (F)
14. I let myself "go" at a party. (T)
15. I consider myself always careful. (F)
16. I change my plans often. (T)
17. I often make people laugh. (T)
18. I like prompt people. (T)
19. I usually notice the furniture arrangements in a strange house. (F)
20. I usually have a ready answer. (T)
21. I don't like to wait for traffic lights to change. (T)
22. I frequently feel "on top of the world." (T)
23. I like work requiring patience and carefulness. (F)
24. I like work involving competition. (T)

25. I remember the names of people I meet. (T)
26. I easily become impatient with people. (T)
27. In watching games, I often yell along with the others. (T)
28. I don't like having my plans changed. (F)
29. I don't like to work with slow people. (T)
30. I'm always on time for social events. (F)
31. I make up my mind easily. (T)
32. I like work in which I must change often from one task to another. (T)
33. I keep a diary regularly. (F)
34. I scan newspapers rather than read them carefully. (T)
35. I like work that has lots of excitement. (T)
36. I like new situations. (T)
37. I have more trouble concentrating than other people seem to have. (T)
38. When I see a train, I wish I were on it. (T)
39. I like to play chess. (F)
40. My interests tend to change quickly. (T)
41. I like to do things on the spur of the moment. (T)
42. I don't like changes. (F)
43. My friends consider me to be happy-go-lucky. (T)
44. I like a great deal of variety in my work. (T)
45. I like being where there is something going on all the time. (T)

EIS

1. Would you describe yourself as an ambitious person? (Y)
2. Do you take things easy when working on a job? (N)
3. Do you like to lie in bed at the weekends? (N)
4. Do you tend to daydream about success rather than work for it? (N)
5. Are you inclined to take life as it comes without much planning? (N)
6. Do you often wish you were a gentleman or a gentlewoman of leisure? (N)
7. Do you tend to put things off to do tomorrow? (N)
8. Are you more happy to read of the successes of others than do the work of making yourself a success? (N)
9. Have you often failed to complete a job or course you started? (N)
10. Are you satisfied if you just get through an examination? (N)
11. Will days go by without your having done a thing? (N)
12. Are you more inclined to do something for someone else rather than get on with your own work? (N)
13. Do you or did you do little work for examinations? (N)
14. Would you describe yourself as being lazy? (N)
15. Are you or were you able to take quite easily being jilted by a boy friend or girl friend? (N)
16. Do you easily forget your work when you go home? (N)
17. Do you sometimes wish you could get away on your own to work on an important task? (Y)

18. Do you often find it difficult to keep on top of your work? (Y)
19. Do the great achievements of others sometimes make you feel small? (Y)
20. Do you often discuss your work with your relatives or friends? (Y)
21. Are you happiest when doing a job as part of a team? (Y)
22. Do you become bored when on holiday? (Y)
23. Do you grow excited when telling someone about the work you are doing? (Y)
24. Are you often quite envious of the success of others? (Y)
25. Are you often amazed at how hard others work? (Y)
26. Do you often compare how you can do something with how well others can do it? (Y)
27. Have you at any time tried to model your life on that of a successful person? (Y)
28. Do you readily forget work when you are on holiday? (N)
29. Are you very interested in the lives of successful people? (Y)
30. Have you become very anxious at the thought of a difficult task you are about to undertake? (Y)
31. Are you influenced by those around you in the amount of work you do? (Y)
32. Do you often find it difficult to sleep because of excitement over some job you are doing? (Y)
33. Do you often find it difficult to concentrate on what a person is saying because you are thinking about some work you have been doing? (Y)
34. Are you often awed in the presence of very successful people? (Y)

[1]Items are keyed in the direction of high scores on original tests, ie, high anxiety, high need for approval, internal control of reinforcement, high impulsivity, high ego-involvement.

Although EPPS and ACL test protocols were collected from subjects in many of the studies, they were not used in test construction. There is limited evidence of the validity of the subscales derivable from these tests. Consequently, identification of the variable(s) measured by each item would have been questionable. Selecting items on a pure empirical basis would have led to mostly artifactual results.

In Study III the Heineman anxiety scale was substituted for the MAS, and a different locus of control test was used. Neither of these was used in test construction, particularly since the group chosen for item analysis was one from Study IV. For similar reasons, the Nowlis Mood-Adjective Check List was not used in test construction. Neither the ISB nor the specially-devised sentence-completion test was used for test construction, since they were not intended for this purpose and do not have objectively-scoreable items.

TABLE XV

ITEMS RETAINED AFTER ITEM ANALYSES,[1] KEYED AS INDICATED BY ITEM ANALYSIS

Reaction-Time Battery:

MAS	M-C SD	LC	IS	EIS
2 (F)*	6 (F)*	1 (b)	3 (F)*	1 (N)*
3 (T)*	7 (F)*	2 (b)*	7 (F)	3 (N)
6 (T)	9 (T)*	3 (a)	11 (F)*	4 (N)
8 (F)*	10 (T)	6 (b)*	15 (T)*	7 (Y)*
9 (T)*	17 (T)	9 (b)	17 (T)	9 (Y)*
10 (F)*	18 (T)*	14 (b)*	21 (F)*	11 (Y)*
12 (T)*	21 (T)	19 (a)	29 (F)*	12 (N)
13 (F)*	25 (T)	21 (a)	30 (T)*	13 (N)
18 (F)	28 (F)*	23 (a)	39 (F)	14 (Y)*
22 (T)	30 (F)*	26 (a)		24 (N)*
23 (T)		27 (a)		27 (Y)
24 (F)*		33 (a)*		28 (Y)*
25 (F)*		35 (b)*		30 (Y)
28 (T)		36 (b)		33 (Y)
34 (F)*		48 (a)		34 (N)*
39 (T)		51 (a)		
41 (T)		54 (b)		
43 (F)*		57 (a)		
44 (F)*		60 (b)		
46 (F)*				

TABLE XV (CONTINUED)

Finger-Retraction Conditioning Battery:

MAS	M-C SD	LC	LC (continued)
1 (T)*	4 (T)	2 (b)*	36 (a)*
3 (T)*	7 (F)*	5 (a)*	37 (a)*
8 (T)	9 (F)	7 (a)	38 (b)
9 (T)*	10 (F)	9 (b)	39 (a)
10 (F)*	13 (T)	10 (a)*	40 (b)
11 (F)*	14 (F)	11 (a)	42 (b)*
16 (T)	17 (T)	12 (b)	43 (b)*
18 (T)*	19 (F)	13 (b)*	45 (a)*
25 (F)*	21 (F)*	14 (b)*	46 (b)*
26 (F)*	23 (F)	16 (b)*	48 (b)*
28 (F)*	24 (T)	17 (a)	49 (a)
29 (F)	25 (T)	18 (b)	50 (b)
31 (F)*	26 (T)	20 (b)	51 (b)*
33 (F)*	28 (F)	22 (b)	52 (b)*
34 (F)*		24 (b)*	53 (b)*
39 (T)		25 (a)*	54 (a)*
40 (T)		29 (a)*	57 (a)
41 (F)		30 (b)*	59 (b)
45 (T)		32 (a)	
47 (T)		35 (b)*	
		36 (a)*	
		37 (a)*	
		38 (b)	
		39 (a)	
		40 (b)	
		42 (b)*	
		43 (b)*	

TABLE XV (CONTINUED)

Simple Verbal Paired-Associates Learning Battery:

MAS	M-C SD	LC	IS	EIS
1 (T)*	2 (F)*	3 (a)	3 (T)	1 (Y)
3 (F)	4 (F)*	4 (a)	5 (F)	3 (N)
5 (F)*	8 (F)*	5 (b)	13 (F)	4 (N)
11 (F)*	10 (F)	6 (b)*	15 (T)	5 (N)
15 (F)	14 (T)*	10 (b)	19 (F)	7 (N)
22 (F)*	17 (F)*	12 (b)	20 (T)	10 (N)
27 (F)*	20 (F)*	16 (a)	22 (T)	11 (N)
28 (F)*	21 (F)*	19 (a)	24 (T)	14 (N)
29 (T)*	27 (F)*	24 (a)	32 (T)	15 (Y)
31 (F)*	28 (F)	25 (b)	38 (F)	17 (Y)
37 (F)*	29 (F)*	34 (b)	40 (F)	24 (Y)
38 (F)	30 (T)*	35 (b)*		25 (Y)
39 (F)*	32 (F)	37 (b)		26 (Y)
42 (F)*	33 (F)*	42 (a)		28 (N)
46 (F)*		44 (a)		30 (N)
47 (F)*		47 (a)		33 (Y)
		49 (a)		
		54 (b)		

TABLE XV (CONTINUED)

Concept-Formation Task Battery:

MAS	M-C SD	LC	IS	EIS
1 (F)	1 (T)	1 (a)*	3 (F)*	2 (N)
5 (T)	4 (T)	6 (a)	8 (F)*	3 (Y)*
8 (T)	10 (T)*	9 (b)	9 (T)*	4 (N)
9 (F)	12 (F)	13 (b)*	13 (T)*	6 (N)
11 (T)	13 (F)*	14 (a)	14 (T)	8 (Y)*
14 (T)	22 (T)*	15 (a)	16 (F)*	9 (N)
17 (T)	23 (F)	17 (b)*	21 (T)	10 (Y)*
18 (F)	28 (T)*	19 (b)*	24 (F)*	12 (N)
22 (T)	30 (T)*	26 (a)	25 (F)*	15 (N)
24 (T)	33 (F)*	33 (a)*	26 (F)*	16 (Y)*
26 (T)		37 (a)*	31 (T)	20 (N)*
27 (T)		40 (b)	42 (T)*	21 (Y)
		44 (b)*		27 (Y)
		45 (a)*		30 (Y)
		48 (b)*		32 (Y)
		51 (b)*		34 (Y)
		52 (b)*		
		60 (b)		

[1]Items are keyed to predict criterion, ie, speed (not latency) on reaction-time task, frequency and latency of CRs on finger-retraction task, number of correct responses on verbal learning task, number of errors on concept-formation task. Asterisks indicate item was keyed opposite to keying on original test. If original test score was originally inversely related to criterion, most items will be keyed opposite to original.

Correlations of original test scores with criteria occasionally vary from those reported earlier. This is because data were discarded for subjects who did not complete all tests administered to their group.

RESULTS OF CROSS-VALIDATIONS

Cross-validation results can be discussed only in general terms. Interpretation of the results is compromised by several problems. In almost all correlations, the number of subjects involved was quite low. Consequently, the correlations are likely to be unusually unstable, and also attenuated by the joint effect of unreliability and small sample size. In many cases, the same test might be expected to correlate positively with criterion measures in one case but negatively in another. This might be expected when different criteria were used, when different tasks were used, or when the experimental conditions were different. One product of this research was to make abundantly clear that motivational variables have different effects on different aspects of performance and in different conditions. The large number of replications partly compensates for these limitations. Several small-sample replications may give less distorted results than a single large-sample cross-validation. Again, however, the "replications" often turned out to differ in important respects from one another.

Overall, it is clear that the test batteries, in present form, are not yet likely to be of practical use for controlling individual differences in experimental research. That this would be true was more or less expected from the outset.

There are encouraging exceptions, however. For example, the short-form MAS predicted performance even better than the original MAS in the control condition in Study XII, the only real replication of the MAS short form based on the concept-formation task data from Study XI. Since this task seemed to be the only one sensitive to Drive effects, this finding was extremely encouraging. In itself it lent considerable support to the evidence that the task is, in fact, sensitive to Drive differences.

These were also scattered indications that the battery devised for the reaction-time task may have some utility in some conditions. The reaction-time battery, in cross-validations, predicted fairly consistently as expected. In some groups these correlations were quite high. The probable instability of the correlations, because of small sample size, reduces the practical significance of these, as well as other correlations. Nevertheless, they provided encouragement. Even when, in some groups, correlations were opposite to the general trend, this may often have been a reliable peculiarity of the particular experimental conditions involved. For example, differences in order or in shock conditions seemed often, in the reaction-time studies, to reduce or reverse the effects of motivational variables.

The batteries devised for the finger-retraction conditioning and the simple verbal learning task, however, showed no evidence of utility. This in itself was difficult to interpret. In studies with both of these tasks, experimental conditions had profound effects on performance, by themselves or in interaction with the test-measured variables; and in both studies, the replication studies turned out to be anything but replications. The low correlations obtained in cross-validations could have reflected simply that the different conditions and replications were not comparable in terms of effects of motivational variables, as well as in other respects.

In summary, the batteries in present form may have some utility, but the cross-validation results offer little clear evidence of it.

SUMMARY AND RECOMMENDATIONS

This research confirmed, at the same time, that tests of individual differences in motivational variables can be extremely useful and that they are extremely difficult to develop and employ.

Their utility is twofold. First, as expected at the outset, they often allowed control of considerable intersubject variance. This made it possible to isolate effects of experimental conditions on performance when these effects would otherwise have been relegated to the category of nonsignificant results. Second, they often served, dramatically at times, to point out phenomena which would otherwise have been overlooked. These phenomena almost invariably indicated in turn that otherwise obvious interpretations of findings were in error. For example, it was the M-C SD which first demonstrated that incentive effects, rather than Drive effects, predominated in the reaction-time tasks. This in turn led to the realization that incentive effects were paramount in all the tasks used, if not in fact in virtually all laboratory tasks.

Several difficulties were apparent in developing and using these tests. First, it became obvious that an omnibus battery would seldom be useful. Different variables were related to performance on different tasks and in different conditions. Tests of these variables would be useful sometimes but not others. A collection of such tests would include much chaff in any one situation, and perhaps include tests which would lead to opposite (but valid) predictions of performance.

TABLE XVI

RESULTS OF CROSS-VALIDATIONS: CORRELATIONS OF BATTERY SCORES, SEPARATE SHORT-FORM SCORES, AND ORIGINAL TEST SCORES WITH CRITERIA IN EACH EXPERIMENTAL CONDITION (DECIMALS OMITTED).

Study	Condition	Criterion	Original Tests					Short-Form Tests					Battery
			MAS	MC-SD	LC	IS	EIS	MAS	MC SD	LC	IS	EIS	
II	NS	Speed (C)	-38	44	23			-27	31				21
		Errors (C)	11	02	-23			-02	19				-21
	S	Speed (C)	-71	-08	02			18	-01				16
		Errors (C)	41	-50	-25			-44	10				-31
III	NS Sp1	Latency (S)		03					32				
		Latency (C)		44					-04				
		Errors (C)		-08					-43				
	NS Sp2	Latency (S)		-18					-41				
		Latency (C)		-37					01				
		Errors (C)		-02					35				
	NS Ac1	Latency (S)		-20					-22				
		Latency (C)		10					-16				
		Errors (C)		-03					-36				
	NS Ac2	Latency (S)		27					-28				
		Latency (C)		28					00				
		Errors (C)		20					30				
	S Sp1	Latency (S)		44					-20				
		Latency (C)		36					03				
		Errors (C)		46					27				

TABLE XVI (CONTINUED)

Study	Condition	Criterion	Original Tests MAS	Original Tests MC-SD	Original Tests LC	Original Tests IS	Original Tests EIS	Short-Form Tests MAS	Short-Form Tests MC SD	Short-Form Tests LC	Short-Form Tests IS	Short-Form Tests EIS	Battery
	S	Latency (S)		-03					-47				
	Sp2	Latency (C)		-28					-57				
		Errors (C)		-34					-25				
III	S	Latency (S)		23					22				
	Ac1	Latency (C)		36					-13				
		Errors (C)		03					-17				
	S	Latency (S)		-12					06				
	Ac2	Latency (C)		-15					09				
		Errors (C)		10					16				
IV	NS	Latency (S)	-07	-02	05	-71	-43	(-73)	(21)	(23)	(-57)	(-71)	(-77)
	1	Latency (C)	-07	-03	03	-71	-44	(-73)	(20)	(23)	(-57)	(-71)	(-77)
		Errors (C)	-03	49	40	41	41	(41)	(-47)	(31)	(14)	(24)	(40)
	NS	Latency (S)	40	-41	-10	-05	13	-13	01	-20	-15	25	-00
	2	Latency (C)	39	-39	-08	-08	13	-11	01	-18	-16	23	-02
		Errors (C)	40	-16	-57	14	-45	-48	19	04	36	-23	-39
	S	Latency (S)	-00	-03	10	31	-00	11	12	12	37	12	36
	1	Latency (C)	-01	-02	11	31	00	11	12	14	39	14	38
		Errors (C)	-03	-21	-03	02	-22	-20	05	-34	10	43	-03
	S	Latency (S)	27	-12	12	-34	-06	-20	19	-02	-33	31	-36
	2	Latency (C)	28	-13	11	-32	-05	-20	19	-00	-33	28	-35
		Errors (C)	-26	29	34	-06	16	31	02	-07	14	21	43

TABLE XVI (CONTINUED)

Study	Condition	Criterion	Original Tests MAS	MC-SD	LC	IS	EIS	Short-Form Tests MAS	MC-SD	LC	IS	EIS	Battery
V	Cont	Latency (S)	03	-49	17	-02	-05	07	-19	-00	-47	-23	-28
		Latency (C)	19	13	61	-62	38	68	-00	61	-76	54	15
		Errors (C)	02	19	-36	-29	-06	-16	03	-17	25	-22	-22
	S	Latency (S)	50	-78	-56	-65	-34	-68	74	-44	-72	28	-87
		Latency (C)	47	-55	-30	-49	-28	-31	25	-47	-10	14	-52
		Errors (C)	-44	65	64	32	61	73	-82	-02	84	-55	68
	L	Latency (S)	-11	47	27	13	09	26	-28	-13	31	-07	44
	Cont	Latency (C)	49	-51	-75	-29	52	-82	-06	36	24	-41	-73
		Errors (C)	-60	-01	-44	25	15	-44	31	-58	-24	02	-15
	L	Latency (S)	04	04	29	32	-06	74	52	-44	-12	-49	30
	Inc	Latency (C)	-06	-43	21	-00	22	54	74	-47	-30	-12	27
		Errors (C)	-28	61	25	48	-16	04	-67	42	49	-16	42
VI	LS	Freq	-11	18	-30			(-54)	(-64)	(-50)			(-50)
	CC	Lat	11	-08	30			(56)	(64)	(51)			(51)
	HS	Freq	-02	-35	18			-01	28	20			16
	CC	Lat	-05	21	03			01	-21	-08			-02
	LS	Freq	10	12	-14			03	-02	-22			-10
	AC	Lat	-11	-12	21			-07	-01	28			08
	HS	Freq	-19	-00	02			-14	-12	-09			-17
	AC	Lat	01	20	-04			27	08	10			21

TABLE XVI (CONTINUED)

Study	Condition	Criterion	Original Tests MAS	Original Tests MC-SD	Original Tests LC	Original Tests IS	Original Tests EIS	Short-Form Tests MAS	Short-Form Tests MC-SD	Short-Form Tests LC	Short-Form Tests IS	Short-Form Tests EIS	Battery
VIII	LS	Freq	03	17	06			-05	08	06			07
	CC	Lat	-20	17	08			07	05	-04			03
IX	NS N	Correct	-39	24	-43			-28	-39	-03			-30
	NS A	Correct	16	-23	10			02	05	-05			10
	NS D	Correct	61	-01	10			15	-17	41			27
	S N	Correct	-25	50	-24			-16	-17	-05			-20
	S A	Correct	-08	-05	-30			-02	03	18			09
	S D	Correct	44	-47	29			40	39	-12			48
X	NS N	Correct	36	-43	-12	-15	43	(02)	(-25)	(39)	(-02)	(07)	(75)
	NS A	Correct	-24	39	-24	17	04	-21	-12	-20	33	01	11
	NS D	Correct	26	21	-29	-67	-63	-66	08	-33	24	24	-72
	S N	Correct	-22	29	34	04	-53	51	-53	42	35	03	02
	S A	Correct	-11	10	07	-13	09	-20	17	17	00	41	01
	S D	Correct	16	04	26	-16	10	22	00	-28	-03	11	18

TABLE XVI (CONTINUED)

Study	Condition	Criterion	Original Tests MAS	MC-SD	LC	IS	EIS	Short-Form Tests MAS	MC-SD	LC	IS	EIS	Battery
XI	C1 W1	Errors C	24	08	-75	10	20	(02)	(16)	(23)	(-16)	(11)	(07)
	C1 F1	Errors C	-27	10	60	-18	21	(-13)	(-03)	(84)	(20)	(23)	(04)
	W1 W1	Errors W	31	24	-05	60	28	(73)	(-70)	(25)	(77)	(72)	(86)
	W1 F1	Errors W	33	30	55	50	13	(26)	(42)	(53)	(53)	(56)	(62)
XII	Cont	Errors C	23					61					
	S	Errors C	-21					-05					
	S Inc	Errors C	-42					-37					

Explanations: Parenthesized correlations are those in groups on which item analyses were based; they are, consequently, usually spuriously high. Where correlations are not reported, tests were not given to subjects in that group. Correlations were computed only from subjects with scores on all tests administered to that group. Sample sizes vary; most were approximately 15.

Abbreviations of Conditions: S: Shock. Sp: Speed Instructions. Ac: Accuracy Instructions. 1: Order 1 (simple task first, followed by complex task). 2: Order 2 (complex task first followed by simple task). Cont: Control condition. L Cont: Light control condition (no incentive). L Inc: Light Incentive condition. LS: Low Shock. HS: High Shock. CC: Classical condition. AC: Avoidance condition. N: Neutral approval condition. A: Approving condition. D: Disapproving condition. C 1, W 1: Color task taken first; Word practice task taken first. C 1, F 1: Color task taken first, Form practice task taken first.

TABLE XVI (CONTINUED)

W 1, W 1: Word task taken first, Word practice task taken first. W 1,. F 1: Word task taken first, Form practice task taken first. S Inc: Shock Incentive condition.

Abbreviations of Criteria:

Speed: Speed scores (reciprocal transformation of latency scores), summed across all trials. Errors: total number of errors. (C): Complex task. (S): Simple task. Latency: latency scores summed across all trials. Freq: total number of CRs. Lat: mean latency of CRs. Correct: total number of correct responses.

Furthermore it became obvious that test-measured variables could be expected to have different, and often opposite, relationships to performance in different conditions. Finally it was apparent that test development could progress no further than the development of tasks which are sensitive to the variables measured by the tests; and, in the case of tests of Drive differences, appropriate tasks are very difficult to devise.

As for the contribution of this research to understanding of motivation phenomena, considerable replication is needed. No clear evidence was obtained bearing on the relative validity of Drive theory and Arousal theory propositions. The findings with the concept-formation task were equally consistent with both positions. The results with the finger-retraction task appeared relevant only until the failure to replicate them cast doubt on their implications. In the other tasks, both Drive and Arousal phenomena were obscured.

The major finding of the research was the overweaning influence of incentive variables on performance, even when this influence was least suspected in advance. Noxious stimuli, and the "social" incentive of the experimenter's behavior, had profound effects. Also potent were individual differences in incentive-relevant variables, notably need for approval and locus of control. These latter variables had been found in previous research to be related to differences in laboratory-task behavior, but only on tasks where they were specifically suspected to have some effect.

In summary, more research is necessary before tests can be developed which will have major practical utility in laboratory studies; but there is every indication that such research will be successful and profitable. The first effort should be to devise sensitive and appropriate criterion tasks, especially for developing a test of individual differences in chronic Drive or Arousal level. The concept-formation task might prove to be adequate for this purpose, but more evidence of its adequacy is needed. No other available task yet seems to be adequate at all. When tasks are developed, tests should be developed from the largest possible item pools, with large numbers of subjects, to produce as powerful a test as possible. Finally, many tasks should be used, in many experimental conditions, to determine where and how the tests will be useful.

REFERENCES

1. Cannon, W. B., The Wisdom of the Body, Norton, New York, 1932.

2. Bernard, C., Lecons sur les Proprietes Physiologiques et les Alterations Pathologiques des Liquides de l'Organisme, Balliere, Paris, 1859.

3. Campbell, B. A., and F. D. Sheffield, "Relation of Random Activity to Food Deprivation," Journal of Comparative and Physiological Psychology, Vol 46, pp 320-322, 1953.

4. Young, P. T., Motivation and Emotion, Wiley, New York, 1961.

5. Pribram, K. H., "Implications for Systematic Studies of Behavior," Chap 39, Electrical Stimulation of the Brain, University of Texas Press, Austin, 1961.

6. McClelland, D. C., J. W. Atkinson, R. A. Clark, and E. L. Lowell, The Achievement Motive, Appleton-Century-Crofts, New York, 1953.

7. Schachter, S., The Psychology of Affiliation, Stanford University Press, Stanford, 1959.

8. Olds, J., "Physiological Mechanisms in Reward," in Nebraska Symposium on Motivation, University of Nebraska Press, Lincoln, 1955.

9. Miller, N. E., "Implications for Theories of Reinforcement," Chap 40, Electrical Stimulation of the Brain, University of Texas Press, Austin, 1961.

10. Hull, C. L., Principles of Behavior: an Introduction to Behavior Theory, Appleton-Century-Crofts, New York, 1943.

11. Spence, K. W., Behavior Theory and Learning, Prentice-Hall, New York, 1960.

12. Brown, J. S., The Motivation of Behavior, McGraw-Hill, New York, 1961.

13. Spence, K. W., "A Theory of Emotionally-Based Drive (D) and its Relation to Performance in Simple Learning Situations," American Psychologist, Vol 13, pp 131-141, 1958.

14. Taylor, Janet A., "The Relationship of Anxiety to the Conditioned Eyelid Response," Journal of Experimental Psychology, Vol 41, pp 81-92, 1951.

15. Elias, M. F., Drive and Incentive in Classical and Avoidance Conditioning, Unpublished doctoral dissertation, Purdue University, 1963.

16. Hebb, D. O., "Drives and the C.N.S. (Conceptual Nervous System)," Psychological Review, Vol 62, pp 243-254, 1955.

17. Malmo, R. B., "Anxiety and Behavioral Arousal," Psychological Review, Vol 64, pp 276-287, 1957.

18. Duffy, E., "The Psychological Significance of 'Arousal' or 'Activation'," Psychological Review, Vol 64, pp 265-275, 1957.

19. Yerkes, R. M., and J. D. Dodson, "The Relation of the Strength of Stimulus to Rapidity of Habit Formation," Journal of Comparative Anatomy, Vol 18, pp 459-482, 1908.

20. Broen, W. E., and L. H. Storms, "A Reaction-Potential Ceiling and Response Decrements in Complex Situations," Psychological Review, Vol 68, pp 405-415, 1961.

21. Murray, H. A., Explorations in Personality, Oxford University Press, New York, 1938.

22. Edwards, A. L., Manual for the Edwards Personal Preference Schedule, Psychological Corporation, New York, 1953.

23. French, Elizabeth G., "Some Characteristics of Achievement Motivation," Journal of Experimental Psychology, Vol 50, pp 232-236, 1955.

24. Atkinson, J. W. and W. R. Reitman, "Performance as a Function of Motive Strength and Expectancy of Goal-Attainment," Journal of Abnormal and Social Psychology, Vol 53, pp 361-366, 1956.

25. Rotter, J. B., Social Learning and Clinical Psychology, Prentice-Hall, New York, 1954.

26. Uhlinger, Carolyn A. and M. W. Stephens, "Relation of Achievement Motivation to Academic Achievement in Students of Superior Ability," Journal of Educational Psychology, Vol 51, pp 259-266, 1960.

27. Barrett, H. B., Sensory Interaction: a Reinvestigation, Unpublished master's thesis, Purdue University, 1964.

28. Campbell, D. T., and D. W. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, Vol 56, pp 81-105, 1959.

29. Garner, W. R., H. W. Hake, and C. W. Eriksen, "Operationism and the Concept of Perception," Psychological Review, Vol 63, pp 149-159, 1956.

30. Kimble, G. A., Hilgard and Marquis' Conditioning and Learning, Second Edition, Appleton-Century-Crofts, New York, 1961.

31. Grant, D. A., "Classical and Operant Conditioning," Chap 1, Categories of Human Learning, Academic Press, New York, 1964.

32. Bitterman, M. E., "Comment," Psychonomic Science, Vol 1, p 94, 1964.

33. Wald, A., Sequential Analysis, Wiley, New York, 1947.

34. Crowne, D. P., and D. Marlowe, The Approval Motive, Wiley, New York, 1964.

35. Rotter, J. B., M. Seeman, and S. Liverant, "Internal vs External Control of Reinforcement, a Major Variable in Behavior Theory," pp 473-517, Decisions, Values, and Groups, Vol 2, Macmillan, New York, 1962.

36. Costello, C. G., Two Personality Scales of Ego-Involvement, Unpublished preprint, Regina General Hospital, Saskatchewan, 1964.

37. Barratt, E. S., "Anxiety and Impulsiveness Related to Psychomotor Efficiency," Perceptual and Motor Skills, Vol 9, pp 191-198, 1959.

38. Elias, M. F., The Effect of Experimentally Induced Drive on Simple, Disjunctive and Cross-Disjunctive Reaction-Time, Unpublished master's thesis, Purdue University, 1961.

39. Chiles, W. D., "Effects of Shock-Induced Stress on Verbal Performance," Journal of Experimental Psychology, Vol 56, pp 159-165, 1958.

40. Ramond, C. H., "Anxiety and Task as Determiners of Verbal Performance," Journal of Experimental Psychology, Vol 46, pp 120-124, 1953.

41. Sheflin, J. A., The Role of Incentives in a Classical Drive Experimental Situation, Unpublished master's thesis, Purdue University, 1965.

42. Schaeuble, J. L., Individual Differences in Verbal Learning: Anxiety, Need for Approval, and Locus of Control, Unpublished master's thesis, Purdue University, 1964.