Stress, Manifest Anxiety, and Clinical Judgment

Joseph F. Pribyl

Loyola University Chicago

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1967 Joseph F. Pribyl
Stress, Manifest Anxiety, and Clinical Judgment

by

Joseph F. Pribyl

A Dissertation Submitted to the Faculty of the Graduate School of Loyola University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

June

1967
Vita

Joseph F. Pribyl was born in Chicago, Illinois on October 29, 1940.

He was graduated from St. Ignatius High School, Chicago, in June, 1958. He received the degree of Bachelor of Science in the Natural Sciences from Loyola University, Chicago, in June, 1962.

Mr. Pribyl began his graduate studies in the Department of Psychology of Loyola University in September, 1962. Until June, 1964, he worked as a research and teaching assistant in the Psychology Department at Loyola. As a Psychologist Trainee at the Illinois Youth Commission Reception Center in Joliet, Illinois from June, 1964, through June, 1965, he completed his clerkship in clinical psychology. In June, 1965, Loyola University conferred the Masters of Arts degree in Psychology upon Mr. Pribyl. As a United States Public Health Fellow, Mr. Pribyl completed his clinical psychology internship at the Neuropsychiatric Institute, Chicago, from July, 1965, through June, 1966. From July, 1966 to the present, he conducted his dissertation research while on a fellowship from the Department of Mental Health of the State of Illinois.
Acknowledgments

The author wishes to express his gratitude to Dr. Ronald E. Walker for his help and encouragement during the preparation of this dissertation. He is also indebted to the members of his committee, Dr. Robert C. Nicolay, Dr. Patricia M. Barger, Dr. Patrick R. Laughlin, and Dr. Homer Johnson, for their suggestions and advice. An expression of thanks is also extended to Dr. William A. Hunt of Northwestern University for his helpful comments and ideas. The stipend received by the author during the preparation of this dissertation from an Employment-Education Grant of the Mental Health Department of the State of Illinois is much appreciated. Last, but not least, the author wishes to thank his wife, Mary, for her suggestions and assistance in completing various aspects of this dissertation.
Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Statement of the Problem and Review of Related Literature</td>
<td>1</td>
</tr>
<tr>
<td>II. Method</td>
<td>17</td>
</tr>
<tr>
<td>III. Results</td>
<td>24</td>
</tr>
<tr>
<td>IV. Discussion</td>
<td>46</td>
</tr>
<tr>
<td>V. Summary</td>
<td>55</td>
</tr>
<tr>
<td>References</td>
<td>57</td>
</tr>
<tr>
<td>Appendix A--Equipment Used in Stress Manipulation</td>
<td>61</td>
</tr>
<tr>
<td>Appendix B--Pre-test Rating Scales</td>
<td>62</td>
</tr>
<tr>
<td>Appendix C--Post-test Questionnaire</td>
<td>63</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mean Interjudge Reliabilities for 20 Subjects Under Each Experimental Condition for all Trials</td>
<td>25</td>
</tr>
<tr>
<td>2. Standard Deviations of the Mean Interjudge Reliabilities for Each Experimental Group for all Trials</td>
<td>26</td>
</tr>
<tr>
<td>3. Analysis of Variance of the Mean $z'$ Values of the interjudge Reliabilities for the Six Experimental Groups Over Nine Trials</td>
<td>28</td>
</tr>
<tr>
<td>4. Mean Interjudge Reliabilities for all of the Stress Subjects and for all of the Neutral Subjects Over Nine Trials</td>
<td>34</td>
</tr>
<tr>
<td>5. Summary of Stress Subjects' Responses to the Pre-test Rating Scales and to the Post-test Question &quot;How Did you Feel About Being in a Situation in Which You Might Receive a Shock?&quot;</td>
<td>43</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Main effect of anxiety upon mean interjudge reliability</td>
<td>29</td>
</tr>
<tr>
<td>2. Mean interjudge reliability of all experimental subjects over trials</td>
<td>31</td>
</tr>
<tr>
<td>3. Interaction of stress and trials</td>
<td>32</td>
</tr>
<tr>
<td>4. Interaction of anxiety and trials</td>
<td>35</td>
</tr>
<tr>
<td>5. Trial-by-trial means of the High Anxious and Low Anxious subjects tested under neutral instructions</td>
<td>37</td>
</tr>
<tr>
<td>6. Trial-by-trial means of High Anxious and Low Anxious subjects tested under stress instructions</td>
<td>38</td>
</tr>
<tr>
<td>7. Trial-by-trial means of High Anxious Stress subjects and High Anxious Neutral subjects</td>
<td>39</td>
</tr>
<tr>
<td>8. Trial-by-trial means of Low Anxious Stress subjects and Low Anxious Neutral subjects</td>
<td>40</td>
</tr>
</tbody>
</table>
Chapter I

Statement of the Problem and Review of Related Literature

Historical Background. Until Meehl (1954) presented his discussion of clinical judgment as analogous to actuarial prediction, many practicing clinicians felt that clinical judgment was such an individual phenomenon that it was outside the realm of research. Meehl suggested that a clinician uses a finite number of facts and "rules of thumb" which he puts together in different combinations of importance to make predictions. The operations which a clinician used in making predictions could be formalized in actuarial tables and weighting formulas which a clerical worker could use in making very accurate predictions. The advantage of the actuarial method over the clinical method would be due to the use of empirically determined optimal weights for different classes of facts rather than vague "rules of thumb."

Hunt and Jones (1962) admit that the actuarial method provides potential for very accurate clinical predictions but point out that high accuracy will not be realized until highly refined actuarial tables are developed. Both Meehl (1954) and Hunt and Jones (1962) realize that the actuarial methods, if they are ever widely used, will be used in conjunction with the clinician creatively producing hypotheses.

Sarbin, Taft, and Bailey (1960) produced a complex clinical judgment model based upon syllogistic reasoning. Hoffman (1960) presented a
mathematical model of clinical judgment based upon information theory. While both approaches have theoretical importance, neither model has resulted in testable hypotheses. Sarbin et al. gave an extensive rational analysis of clinical judgment, but presented few testable hypotheses. Hoffman has not sufficiently identified the cues or inputs so important to information theory to make his hypotheses testable. (Hunt & Jones, 1962).

Bieri, Atkins, Briar, Leaman, Miller, and Tripodi (1966) reviewed much of the literature of clinical and social judgment and compared it with the findings of classical psychophysics. They carefully applied information theory models to the findings of social and clinical judgment studies and psychophysical research. Since situational variables have not been extensively explored in previous research, Bieri et al. developed a research framework to provide an orientation for future research relating clinical judgment and situational variables.

While the problems of developing and testing new clinical judgment models are complex, Hunt (1959) felt that currently used clinical methods could be improved through research. Underwood (1957) contends that scientific investigation is only possible with phenomena that are reliably (repeatedly) obtained. The unique character of the clinical situation would seem to make research impossible. By focusing research efforts on the clinician making repeated judgments, Hunt (1959) felt that the commonalities of the judgmental situation could be used as a
basis for rigorous scientific investigation. Variables that influence agreement among several judges could be studied by comparing judgmental performance in identical, or at least similar, situations.

The situation in which several clinicians are asked to make repeated judgments on the clinical material is analogous to the paradigm of classical psychophysics (Hunt, 1959). It is Hunt's hope that clinical judgment can be shown to be one of several phenomena embodied in the general category of judgment. If this is the case, much of the literature pertaining to psychophysical judgment can be brought to bear on the problem of clinical judgment.

Using the psychophysical model as a base, Hunt and Arnoff (1955, 1956) demonstrated that clinical judgment is reliable as measured by interjudge agreement. Other workers (Campbell, Hunt, & Lewis, 1957; Campbell, Lewis, & Hunt, 1958) have shown that the context effects well known to classical psychophysics (Beeb-Center, 1929; Helson, 1947; 1948; Hunt, 1941; Hunt & Volkman, 1937; Johnson, 1955) are also found in clinical judgment.

Learning has been shown to be important in classical psychophysics. Helson (1947; 1948) demonstrated that Ss' previous acquaintance with similar stimuli shifts the Ss' adaptation level (a phenomenon in which previously perceived stimuli influence perception of subsequent stimuli). Experienced clinicians should be better able to make meaningful clinical judgments because of the clinicians' past experience with a wider range
of stimuli. Several investigators (Grigg, 1958; Hunt, Jones, & Hunt, 1957; Jones, 1957; Cline, 1955) have confirmed the above.

In efforts to relate clinical judgment to other areas of psychological research, Hunt and his colleagues began to investigate the relationship between learning theory and clinical judgment. In doing so, Hunt and Jones (1962) hoped to obtain a broader experimental base for clinical judgment. Gibson (1953) reviewed a number of studies which showed that absolute judgments made with the method of single stimuli improve even when there is only practice without knowledge of results. Ammons (1956) reviewed many studies which dealt with different types of judgments and perceptual-motor performances. He concluded that in general, learning is faster and reaches a higher level with knowledge of results and that the more specific the knowledge of results, the more rapid is the improvement in performance.

After considering literature which related judgment and knowledge of results, Blumberg (1961) conducted an experiment in which he gave naive judges three levels of knowledge of results. The levels were no feedback, general feedback, (judges were told whether their rating was right or wrong), and specific feedback (judges were told the standard rating after they gave their rating). He expected the judges to increase the reliability, validity, and rapidity of repeated clinical judgments without feedback, but also expected the reliability, validity, and rapidity of judgments to increase as feedback became more specific. In addition, he expected transfer of training when new clinical stimuli were judged. Having judges
rate the vocabulary test responses of hospitalized schizophrenics on a 7-point scale of exhibited disorganization, he found that the three levels of feedback made no differences in the rapidity of the judgments and that the reliability and validity of the clinical judgments did not improve in the no feedback condition. The reliability as measured by interjudge agreement and validity of the clinical judgments did improve when the judges received general and specific feedback. The results of the reliability measure and the validity measure were very similar. The hypothesis that there would be transfer of training in all three conditions was not supported.

M. K. Pribyl (1966) used a task similar to Blumberg’s. She compared the effect of no-feedback with specific feedback and found that Ss who received feedback performed less well than Ss who received no feedback on initial trials. On later trials the feedback and nonfeedback Ss performed approximately equally well with the feedback Ss obtaining slight superiority on the final trial. The initially poor showing of the feedback Ss may have been due to Ss’ confusion as to how to use feedback information. M. K. Pribyl’s suggestion seems reasonable in view of Ammons (1956) review in which he cites several studies in which feedback seemed to cause initial confusion in Ss’ performance on more traditional psychophysical tasks.

Manifest Anxiety and Clinical Judgment. In the field of learning much research has grown out of the concept of drive as measured by anxiety scales (Sarason, 1960). Taylor’s Manifest Anxiety Scale (hereafter referred
to as MAS) was the first such scale to receive widespread attention (1951, 1953, 1956). She originally developed the MAS as an operational measure of Hull's drive in an eyelid conditioning experiment (Taylor, 1951). Taylor developed the Hullian based hypothesis that different sources of drive summate in Ss to produce a total effective drive state (D) that sets the strength of the conditioned eyelid response. Assuming that different levels of psychiatrically defined "manifest anxiety" would be related to different levels of generalized drive, Taylor obtained 65 true-false items (from a pool of 200 MMPI items) which 80 per cent of a group of clinical staff members chose as indicative of operationally defined manifest anxiety. The original 65 items were later cut to the 50 that had the highest correlation with the total score (Taylor, 1953).

Taylor (1951) found that Ss who scored high on the MAS (hereafter high anxious Ss will be referred to as HA Ss) were consistently superior to Ss who scored low on the MAS (low anxious Ss hereafter referred to as LA Ss) in the amount of eyelid conditioning. After 20 eyelid conditioning trials the introduction of stress instructions failed to produce statistically significant effects. Taylor interpreted the differential eyelid conditioning obtained for the HA and LA Ss as meaning the drive level of the HA Ss was higher than that of the LA Ss. Hence, the growth curves of the excitatory potentials for the two groups of Ss were different. On the basis of Hull's (1943) postulate that the growth of excitatory potential was dependent upon both habit strength (H) and drive (D), Taylor suggested
that the difference in the growth curves of excitatory potential in the two groups (inferred from differences in the conditioning curves) might be due to changes in both \( D \) and \( H \). If the above postulate were correct, the HA \( S_s \) would have a stronger reaction to the unconditioned stimulus implying that the same physical stimulus had a different psychological value for the HA \( S_s \) and LA \( S_s \). According to Hull's (1943) postulate reward partially determines \( H \). Therefore, the termination of the unconditioned stimulus should produce a greater reduction of \( D \) in the HA \( S_s \) and therefore, increase \( H \).

In the situation in which only one habit is evoked (such as eyelid conditioning) higher drive level (inferred from higher MAS scores) should lead to better performance. For tasks in which there are several habits having differing levels of availability the predictions are more complex. In a complex task two other Hullian (1943) concepts, oscillatory inhibition \( (O) \) and threshold \( (L) \), must be used. Oscillatory inhibition varies from moment to moment such that its distribution for a group of individuals on the same response at a given moment would be approximately normal. In addition, \( O \) plays an inhibitory role, subtracting from excitatory potential such that there are varying levels of momentary excitatory potential. For a given response to occur, the momentary excitatory potential must be higher than the threshold value for that response. The value of \( L \) is assumed to be the same for like habit tendencies evoked in a particular situation. If several response tendencies are competitively available for a particular
task, the one with the highest momentary excitatory potential will take place. Since excitatory potential is dependent upon habit strength, other things being equal, the response tendency with the greatest H and therefore, the greatest excitatory potential has the greatest probability of taking place. Adding the postulate that D effects excitatory potential, when the correct response is weaker (lower H) than one or more competing response tendency, the Ss with higher D will perform less well than Ss with lower D. It is possible that responses having very weak habit strength may gain enough excitatory potential to be above threshold, thus reducing the probability of the correct response in the high D Ss. If the correct response is maximally available, heightened drive would make performance superior for high drive Ss.

Blumberg (1961) found that learning (improvement in the reliability and validity of judgment) took place when Ss had general and specific feedback. Some measures also indicated that different kinds of learning took place even with no feedback. For example, naive Ss repeatedly rating the same clinical stimuli (schizophrenics' vocabulary test responses) over trials did so at about the same reaction time for each level of feedback (none, general, and specific). Hunt and Blumberg (1961) obtained this finding in a replication of the earlier study. If nothing else, the Ss were learning to give their own judgmental responses more rapidly.

In the clinical situation a practitioner usually increases the speed with which he makes clinical evaluations. By combining the drive theory
interpretation of manifest anxiety and clinical judgment, Hunt and his co-workers hoped to shed light upon the process and progress of learning in clinical judgment. In clinical practice the question of interest is: When does a clinician stop making clinical judgments and merely respond to relevant cues with a previously learned judgment? The experimental analogue to the clinical situation would involve naive Ss making repeated clinical judgments. Since clinical judgment is a complex task, HA Ss should initially perform less well than LA Ss. The difference in performance of the HA Ss and LA Ss should shrink with repeated judgments and they should perform equally well. Taylor's interpretation of drive theory (1956) suggests that HA Ss have a greater response probability for competing responses, making incorrect responses more likely. Once the HA Ss establish the correct response, they should perform with shorter latency than LA Ss. The point at which the performance curves for the HA Ss and LA Ss cross, as predicted by drive theory, would be the point at which evaluative judgment stopped and the elicitation of learned verbal responses began.

Hunt and Blumberg (1961) did not find support for the drive theory interpretation of manifest anxiety when HA Ss and LA Ss rated 211 schizophrenics' vocabulary test responses on a 7-point scale of disorganization in different orders over six trials. Of four measures used (latency, number of shifts in judgment, reliability or interjudge agreement, and validity as represented by the agreement of the judge with the standardized
values of the stimuli) the HA Ss and LA Ss were differentiated on only
two (reliability and validity). On both validity and reliability HA Ss
initially performed less well than LA Ss but the learning curves did not
cross as predicted. Instead, the learning curves stayed at about the
same level, placing this particular application of Taylor's drive theory
in doubt.

Hunt and Walker (1963) re-analyzed the Hunt-Blumberg data and repli-
cated their experiment with exactly the same results. In order to check
the possibility that the judges did not have a sufficient number of clini-
cal judgments for their learning curves to cross, Hunt and Walker (1963)
ran another study in which HA and LA judges rated 100 different schizo-
phrenics' vocabulary responses in ten trials of ten responses each.
Again the drive interpretation of manifest anxiety was not supported.
The HA judges were initially less reliable than the LA judges but never
became superior to the LA judges. Hunt and Walker felt that their results
better fit Child's (1954) drive stimulus interpretation of manifest
anxiety. According to Child, the HA individuals had more task irrelevant
responses than the LA individuals. Through learning, the task irrelevant
responses were dissipated at which time the performance decrement of HA
judges disappeared.

Walker, Hunt, and Schwartz (1965) observed that when they conducted
clinical judgment experiments in co-acting, non-interacting group setting
the _Ss seemed less anxious. Along with Spence (1963), they felt that the
number of task irrelevant responses is related to the amount of psychological stress. They predicted that HA and LA judges making clinical judgments in a group setting would perform equally well through many judgments. This prediction was born out in two studies. In another study, (also reported in Walker et al, 1963), the HA judges were initially superior to the LA judges. Pribyl, Walker, and Hunt (1965) replicated the individual and group conditions with HA and LA judges drawn from the same subject pool and run by the same E. They found that the HA and LA judges run in a group performed as expected in that they were equally reliable through all of the clinical judgment trials. In the individual condition, however, the HA and LA judges were also equally reliable throughout the clinical judgment task. This was not expected but might have been due to an E variable or to a difference between the populations of their study and the previous studies.

In a study using different sets of clinical materials, M. K. Pribyl (1966) found that in an individual condition, HA and LA judges were not significantly different in the reliability of their clinical judgments on the initial trial. Her results may have been an artifact of the stimuli she used but they raise questions about the generality of the findings of the previous studies. She also found that learning took place for all judges as they progressed through the clinical judgment task whether they were given knowledge of results or not. One possibility is that the single random order of the set of 100 schizophrenics' vocabulary test responses used in many of the previous studies (Hunt & Walker, 1963;
Pribyl, Walker, & Hunt, 1965; and Walker, Hunt, & Schwartz, 1965,) may have produced an experimental artifact. Inspection of the graph of the results of one of the studies (Pribyl, 1965) and the trial-by-trial means of some of the other studies (Hunt & Walker, 1963; Walker, Hunt, & Schwartz, 1965) suggest that learning did not take place but that inter-trial differences were largely random.

In any area of research it is desirable to exactly replicate previous studies using identical methodologies. This is not too frequently done, however, because researchers often discover ways of increasing the precision and experimental control of their studies. A case in point is M. K. Pribyl's (1966) study. She partially replicated several previous studies (Hunt & Walker; 1963; Walker, Hunt, & Schwartz, 1965; Pribyl, Walker, & Hunt, 1965) but added tighter controls in the form of automatic presentation of the stimuli so that all Ss in any one experimental condition saw the stimuli for the same amount of time. Previous studies allowed inter-subject variability in the time of presentation of the stimulus depending on how quickly an S called out his rating. The question of importance is whether her results are due to E differences, stimulus differences, or procedural differences which eliminate inter-subject variability in exposure.

In a pilot study using random groups of judges, Pribyl found that student judges, who received instructions implying that those who did poorly on the clinical judgment task would be referred for psychiatric
counseling, did not perform differently from a control group which received only neutral task instructions. This was true whether the testing was done individually or in groups. A partial explanation for this failure might be that the stress instructions were ineffective.

Hunt and Walker (1965) have indicated that manifest anxiety has a small ephemeral effect upon clinical judgment performance. The results of previous research relating clinical judgment and manifest anxiety supports the task irrelevant response to manifest anxiety rather than the drive interpretation. Spence (1963) suggested that stress raises drive level. While studies relating stress, manifest anxiety and clinical judgment (Hunt & Walker, 1963; Pribyl, Walker, & Hunt, 1965; Walker, Hunt, & Schwartz, 1965) have not always found an effect due to stress, the stress condition was a mild one. Sarason (1960) reviewed a number of studies which related stress manipulation of Ss of differing anxiety levels to performance on many tasks. He discussed a number of studies in which stress did not have a differential effect on performance of HA and LA Ss. There are also a number of studies in which HA and LA Ss performance was similar under neutral (task oriented) instructions but different under stress instructions.

In the present study, both HA and LA Ss were used in a clinical judgment task similar to that of the Hunt and Walker (1963) study. A modification of the stress procedures and instructions used by Gerard and Rabbie (1961), Rabbie (1963), and Schachter (1959) were used in this
experiment. They threatened Ss with an electric shock and gave fear inducing instructions and then measured the dependent variable of interest while the Ss thought basal measures were being taken. According to rating scales, galvanic skin response, and questionnaires these experimental procedures reliably aroused stress reactions in college Ss. It was expected that with sufficient stress the HA Ss would initially perform less well than LA Ss and with continued clinical judgments the HA Ss will perform better than the LA Ss as predicted by Taylor (1956) and Hunt and Walker (1963).

An assumption underlying this study is that the drive interpretation of manifest anxiety would be supported if the subjects were tested under conditions of heightened drive. Hence, a stress technique was chosen which has previously been shown to arouse stress reactions in an experimental population similar to that used in this study.

Unlike earlier studies which tested only HA and LA subjects, the present study included a middle anxious (MA) group. This was done to further explore the parameters of performance of Ss of varying anxiety levels as suggested by Sarason (1960). The assumption was made that MA Ss would perform at a level midway between that of HA and LA subjects.

The experimental design for this study thus involves three levels of an organismic variable (HA, MA, LA), a two level treatment variable (stress, neutral), and repeated measures (nine trials). The hypotheses are presented in terms of this factorial analysis of variance design.
Hypothesis One. Anxiety. The mean interjudge reliability of HA judges averaged over all trials of stimuli and both conditions will be less than that of LA judges averaged over all trials of stimuli and both conditions in absolute values but the differences will not be statistically significant. The overall average of the mean interjudge reliabilities of the MA judges will be between that of the HA and LA judges.

Hypothesis Two. Stress. The mean interjudge reliability of all Ss judging under stress instructions will be significantly different from that of all Ss judging under neutral instructions. This hypothesis is based on Schachter's (1959) finding that with the use of his stress instructions there is a difference in performance under stress and non-stress conditions.

Hypothesis Three. Trials. There will be a significant upward linear trend over trials in the average of all Ss' mean interjudge reliabilities. This hypothesis is based on Ammon's (1956) contention that in practice human Ss have some internal feedback on how well they are doing and are, therefore, able to improve performance.

Hypothesis Four. Anxiety x Trials. On the first trial the mean interjudge reliabilities of the HA, MA, and LA judges are expected to be significantly different, with the HA judges lowest and that of the MA and LA judges higher in that order. After the first trial, the trial-by-trial means of the three groups of judges are expected to become much alike in values and then separate slightly with the HA judges highest followed by the MA and then the LA judges. Because only half the Ss
receive stress instructions, the differences among the HA, MA, and LA means on later sets are likely to be very small; and therefore, anxiety by trials interaction is expected to be non-significant or only minimally significant.

**Hypothesis Five. Anxiety x Stress.** A significant interaction is expected, as significant differences among the HA, MA, and LA judges' mean interjudge reliabilities are expected in conjunction with stress and neutral instructions.

**Hypothesis Six. Stress x Trials.** A non-significant interaction is expected. The curve of the trial-by-trial means of all the Ss judging under neutral instructions will start at a higher level and progress upward. The curve of the trial-by-trial means of all the Ss judging under stress instructions is expected to begin at a lower level and, if stress is relatively constant as expected, the curve will rise but will not attain the level of the non-stress group.

**Hypothesis Seven. Anxiety x Stress x Trials.** A significant interaction is expected. Under neutral instructions, the results are expected to replicate the findings of Pribyl (1966) that the learning curves of HA and LA Ss are substantially similar. Under stress instructions, the original Hunt and Walker (1963) hypothesis is expected to be confirmed. HA Ss will start out lowest and will improve faster than MA or LA judges. By the last trial HA Ss will be most reliable, followed by the MA Ss and then the LA Ss.
Chapter II

Method

Subjects. Freshman male general psychology students who were drawn from the subject pool maintained at the Lake Shore Campus of Loyola University participated in the present experiment. Only freshmen served because it was felt that upperclassmen might not be as naive about the experimental instructions in the stress condition. Only males were used to eliminate possible contamination of the experimental results due to sex-related differences in performance of the experimental task. The Ss who served were those who scored in the highest 20 per cent, the middle 20 per cent, and the lowest 20 per cent of the distribution of MAS scores of the freshman male general psychology students. The MAS was routinely administered to all the general psychology students as a classroom exercise during the first two weeks of class. The range of MAS scores for HA Ss was from 23 to 40, for MA Ss the range was from 10 to 16, and for the LA Ss the range was from 1 to 7.

The names of the selected students were placed alphabetically on the front of a folder with appointment times inside. Instructions on the folder indicated that the subjects had been selected randomly. No student had to participate in this particular experiment but all general psychology students were required to participate in five one-hour experiments.

So that the E would not know to which anxiety level a particular
subject belonged, an assistant took the student's name off of the MAS answer sheets and alphabetized them. Departmental assistants administered the MAS to all students, thus making it unlikely that the Ss would associate the MAS with the experimenter.

Twenty Ss from each of the three anxiety levels rated the schizophrenics' vocabulary test responses under neutral instructions. Separate groups of twenty Ss who scored in each of the three anxiety levels rated the same set of schizophrenics' vocabulary test responses with the addition of stress instructions. The Ss were initially assigned to neutral or stress conditions in blocks of ten, the subjects being assigned to a block in the order in which they signed up for the experiment. The order of the stress and neutral blocks were determined through the use of a random number table. The anxiety level of Ss was kept from the E by having an assistant monitor which Ss had been tested under which condition.

After about two-thirds of the Ss had been tested, one of the general psychology teachers unknowingly told his class about the stress technique used in the experiment. Hence Ss from one of the four general psychology classes could no longer be used in the stress condition. Since the potential pool of stress Ss was drastically reduced, an assistant assigned Ss to each experimental condition on a nonrandom basis. Thus, proportionately fewer Ss from the psychology lecture section in question received stress instructions while proportionately more Ss from the other sections were tested under stress conditions. The E, however, was only told under
which condition each S was to be tested. He did not know to which anxiety level an S belonged.

Stimuli. The stimuli used in this study were 90 of the 100 used in a previous series of studies which investigated the relation of clinical judgment and anxiety (Hunt & Walker, 1963; Pribyl, Walker, & Hunt, 1965; Walker, Hunt, & Schwartz, 1965). All are Wechsler Adult Intelligence Scale (WAIS) vocabulary test responses which were rated on a 1 - 7 scale of schizophrenic confusion by a group of experienced clinicians (four or more years of experience). The mean ratings of the expert judges were available for all items. In the previous studies, the 100 items were arranged in 10 sets of 10 stimuli each, with each set containing two items at scale points one, two, and three, and one item at each of the scale points four through seven. The order of items within each set and the order of the sets had been determined randomly for the first of the previous studies (Hunt & Walker, 1963) and had been retained in subsequent studies (Pribyl, Walker & Hunt, 1965; Walker, Hunt, & Schwartz, 1965). In the present study, the particular items in a set was changed to a new random order. The change in the order of the sets permitted some of the characteristics of the sets of stimuli to be assessed independently of the order of presentation.

Procedure. The Ss in the stress condition were greeted in a dry, formal manner by the E wearing a white lab coat. They were ushered into an experimental cubicle containing a chair with wrist electrodes dangling over one arm rest. A series of electrical panels with a number of knobs
and dials was in view as the $S$ entered the room. A tube of electrode jelly, paper towels, and applicator sticks were on a table near the $S$.

Appendix A shows a plate of the $S$s' view of the stress equipment as he entered the experimental room.

The $S$ was seated and the $E$ began thus (a modification of the instructions used by Schachter, 1959, pp. 13-14):

"Allow me to introduce myself. I am Dr. Joseph Pribyl of the Medical School's Neuropsychiatric Research Unit. I have asked you to come today in order to serve as a subject in an experiment concerned with the effects of electric shock.

"What we will ask you to do is very simple. We would like to give you a series of electric shocks. Now, I feel I must be completely honest with you and tell you exactly what you are in for. These shocks will hurt, they will be painful. As you can guess, if, in research of this sort, we're to learn anything at all that will really help humanity, it is necessary that our shocks be intense. What we will do is put an electrode on your arm, hook you into this apparatus (Pribyl points to the electrical-looking gadgetry behind him), give you a series of electric shocks, and take various electrophysiological measures as you perform an experimental task. Again, I do want to be honest with you and tell you that these shocks will be quite painful but, of course, they will do no permanent damage.

"You will be shocked only after you reach a criterion of errors in the experimental task. According to very reliable norms about one in four subjects at this University reach the criterion of errors and are therefore given electric shocks."

Then the instruction for the clinical judgment task were given. They are a modification by M. K. Pribyl (1966, pp. 53-54) of the instructions used by Hunt and Walker (1963).

"We are going to present you with a number of responses made by schizophrenic patients in a mental hospital to vocabulary test
items taken from an intelligence test. One of the ways in which the pathology of schizophrenia may express itself is through disorganized thinking which results in atypical, unusual, or 'abnormal' responses to the items on such a test. The qualitative interpretation by an experienced clinical psychologist of such test responses is one of the bases upon which he may make a clinical or diagnostic interpretation. The extent of the disorganization exhibited in these responses is not uniform. In some of the responses it is minimal and in others it is extreme.

"You are asked to rate these responses on a 7-point scale, from 1 through 7, according to the severity of the disorganization exhibited in the response, with the low end of the scale representing minimal disorganization and the high end of the scale representing maximal disorganization. In making these ratings what we are asking you to do is to judge how 'schizophrenic' each response is. Some responses will seem quite normal; those you would rate '1.' Others will be so disorganized as to require a '7' rating. The majority will fall somewhere in between.

"First you will be shown four sample items to give you an idea of how these responses are rated. The responses you are being asked to judge will be projected on a screen for 10 seconds each and you will be asked to give a rating within that time.

"You will now have a few minutes to look over the samples. You will have an opportunity to ask questions after you have looked them over, but once the experiment starts, you will have to hold all questions until the end of the experiment."

In order to obtain a measure of the stress arousal the E continued.

"Before we begin, I'd like you to tell us how you feel about taking part in this experiment and being shocked. We need this information in order to fully understand your reactions in the shocking apparatus. I ask you therefore to be as honest as possible in answering and describe your feelings as accurately as possible."

The E then asked the S to fill out a questionnaire entitled "How do you feel about being shocked?" A copy of this questionnaire is shown in Appendix B.
The S then studied the four sample items. After the S studied the samples, the E applied electrode jelly to each sides of the S's right wrist and strapped the electrodes on. Then the S's arm was strapped to the arm rest of the chair. The E then "switched on the shocking apparatus."

The Ss in the neutral condition received only the instructions dealing directly with the clinical judgment task. No mention was made of electric shocks, nor were any of the paraphernalia ("shocking apparatus," electrode jelly, etc.) of shocking visible. The E wore a business suit and greeted the S in a more friendly manner.

After the instructions were given, the actual experimental procedure began. Since the E had difficulty hearing some practice S's ratings over the projector blower, the E asked each S to call out his rating in a louder than conversational tone of voice.

The stimuli, which were typed statements presented on 35 mm slides, were projected on a white screen approximately four feet from where the S was seated. The E sat to the right and slightly behind the S. The presentation of the slides was automatic, each slide in the series being presented automatically for ten seconds. During the time that the slide was visible, the S called out his rating and the E recorded it.

After all of the items had been rated, the E asked the S the questions shown in Appendix C. Stress Ss were not told that they could not be shocked. (All of the "shocking apparatus" was inoperative). All Ss were told that after all the data were collected, the E would send a notice
around the general psychology classes stating when and where he would be available to answer questions. They were asked not to discuss the experiment with their friends.

After all the data were collected, the E was available in an office for a total of three days for those Ss who wanted to ask questions. The main purpose of the follow-up was to uncover any undue reaction to the stress instructions. A secondary purpose was to provide another opportunity to discuss the experiment for those subjects who wished to do so.
Chapter III

Results

Each S's rating of each trial of 10 stimuli was correlated with the mean ratings of his group for each trial. Although correlating an S's ratings with a mean which includes his ratings would spuriously raise the correlation (McNemar, 1962), it is felt that this factor was not of great importance with so large a number of subjects. The Pearson product-moment correlations were converted into $z'$ values using a conversion table presented by Blalock (1960). Mean $z'$ values (or mean interjudge reliabilities) were computed for each of the six groups of Ss (HA-stress, MA-stress, LA-stress, HA-neutral, MA-neutral, and LA-neutral) on a trial-by-trial basis. These means are presented on Table 1. The standard deviations of the mean interjudge reliabilities of the six groups on a trial-by-trial basis are presented in Table 2.

The mean interjudge reliability (mean interjudge agreement) was used as a response measure because of interest in determining whether judges respond similarly after they have experience in rating clinical material. The experimental situation is somewhat analogous to determining whether clinicians agree with each other in making diagnoses after they receive training. While a validity measure was available in the clinicians' ratings of the stimuli, it was not used. Previous research (Blumberg,
Table 1

Mean Interjudge Reliabilities for 20 Subjects Under Each Experimental Condition for all Trials. All Means are $z'$ Values

<table>
<thead>
<tr>
<th>Trial</th>
<th>High Anxious</th>
<th></th>
<th>Middle Anxious</th>
<th></th>
<th>Low Anxious</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stress</td>
<td>Neutral</td>
<td>Stress</td>
<td>Neutral</td>
<td>Stress</td>
<td>Neutral</td>
</tr>
<tr>
<td>1</td>
<td>1.017</td>
<td>1.463</td>
<td>1.200</td>
<td>1.362</td>
<td>1.240</td>
<td>1.236</td>
</tr>
<tr>
<td>2</td>
<td>1.422</td>
<td>1.576</td>
<td>1.558</td>
<td>1.579</td>
<td>1.476</td>
<td>1.465</td>
</tr>
<tr>
<td>3</td>
<td>1.381</td>
<td>1.469</td>
<td>1.487</td>
<td>1.405</td>
<td>1.403</td>
<td>1.306</td>
</tr>
<tr>
<td>4</td>
<td>1.323</td>
<td>1.393</td>
<td>1.520</td>
<td>1.532</td>
<td>1.537</td>
<td>1.509</td>
</tr>
<tr>
<td>5</td>
<td>1.179</td>
<td>1.328</td>
<td>1.232</td>
<td>1.330</td>
<td>1.348</td>
<td>1.402</td>
</tr>
<tr>
<td>6</td>
<td>1.373</td>
<td>1.592</td>
<td>1.701</td>
<td>1.455</td>
<td>1.475</td>
<td>1.469</td>
</tr>
<tr>
<td>7</td>
<td>1.118</td>
<td>1.285</td>
<td>1.357</td>
<td>1.281</td>
<td>1.272</td>
<td>1.235</td>
</tr>
<tr>
<td>8</td>
<td>1.419</td>
<td>1.478</td>
<td>1.627</td>
<td>1.458</td>
<td>1.540</td>
<td>1.443</td>
</tr>
<tr>
<td>9</td>
<td>1.653</td>
<td>1.623</td>
<td>1.790</td>
<td>1.710</td>
<td>1.813</td>
<td>1.664</td>
</tr>
</tbody>
</table>
Table 2

Standard Deviations of the Mean Interjudge Reliabilities
for each Experimental Group for all Trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>High Anxious</th>
<th></th>
<th>Middle Anxious</th>
<th></th>
<th>Low Anxious</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stress</td>
<td>Neutal</td>
<td>Stress</td>
<td>Neutal</td>
<td>Stress</td>
<td>Neutal</td>
</tr>
<tr>
<td>1</td>
<td>0.345</td>
<td>0.353</td>
<td>0.325</td>
<td>0.321</td>
<td>0.376</td>
<td>0.586</td>
</tr>
<tr>
<td>2</td>
<td>0.395</td>
<td>0.345</td>
<td>0.327</td>
<td>0.305</td>
<td>0.382</td>
<td>0.457</td>
</tr>
<tr>
<td>3</td>
<td>0.194</td>
<td>0.319</td>
<td>0.265</td>
<td>0.239</td>
<td>0.311</td>
<td>0.333</td>
</tr>
<tr>
<td>4</td>
<td>0.112</td>
<td>0.230</td>
<td>0.142</td>
<td>0.319</td>
<td>0.349</td>
<td>0.436</td>
</tr>
<tr>
<td>5</td>
<td>0.325</td>
<td>0.345</td>
<td>0.395</td>
<td>0.339</td>
<td>0.247</td>
<td>0.355</td>
</tr>
<tr>
<td>6</td>
<td>0.259</td>
<td>0.352</td>
<td>0.127</td>
<td>0.264</td>
<td>0.325</td>
<td>0.404</td>
</tr>
<tr>
<td>7</td>
<td>0.363</td>
<td>0.273</td>
<td>0.354</td>
<td>0.353</td>
<td>0.291</td>
<td>0.298</td>
</tr>
<tr>
<td>8</td>
<td>0.233</td>
<td>0.341</td>
<td>0.316</td>
<td>0.371</td>
<td>0.535</td>
<td>0.512</td>
</tr>
<tr>
<td>9</td>
<td>0.360</td>
<td>0.402</td>
<td>0.424</td>
<td>0.306</td>
<td>0.408</td>
<td>0.376</td>
</tr>
</tbody>
</table>
1961; Hunt & Blumberg, 1961; and Pribyl, Mary K., 1967) has shown that the general psychology students' ratings of confusion of schizophrenics' responses agree very closely with the experienced clinicians' ratings of the stimuli. Furthermore, Mary K. Pribyl (1967) found that the correlations between the mean interjudge reliability and the clinicians' ratings for sets of stimuli range from 0.57 to 0.98 with almost all of the correlations above 0.85. Hence, it may be assumed that the subjects' mean interjudge reliabilities are a relatively meaningful measure of accuracy of ratings.

An analysis of variance was done on the mean interjudge reliabilities, the variables being Stress (two levels), Anxiety (three levels), and Trials (nine). The method of analysis of variance was the one suggested by Edwards (1960) for doing a trend analysis on data involving two or more factors, each at two or more levels, and a trials variable. The analysis of variance results are presented in Table 3.

While there is a slight overall superiority of performance of neutral Ss over that of stress Ss, the main effect of stress is not significant. This is not consistent with the results of some of the previous studies which found that stress was detrimental to performance. The main factor of anxiety was not significant, nor did the overall means fall in the hierarchy expected. Instead, the MA Ss' performance was slightly superior to the LA Ss' performance which in turn was above that of the HA Ss. The graph in Figure 1 shows the relationship of the performance of the Ss of each of the three anxiety levels.
<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress</td>
<td>0.127</td>
<td>1</td>
<td>0.127</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>1.236</td>
<td>2</td>
<td>0.618</td>
<td>1.74</td>
</tr>
<tr>
<td>Stress x Anxiety</td>
<td>2.107</td>
<td>2</td>
<td>1.054</td>
<td>2.96</td>
</tr>
<tr>
<td>Error (a)</td>
<td>40.629</td>
<td>114</td>
<td>0.356</td>
<td></td>
</tr>
<tr>
<td>Trials</td>
<td>20.879</td>
<td>8</td>
<td>2.610</td>
<td>23.51**</td>
</tr>
<tr>
<td>Linear Components</td>
<td>9.101</td>
<td>1</td>
<td>9.101</td>
<td>81.99**</td>
</tr>
<tr>
<td>Quadratic Components</td>
<td>3.061</td>
<td>1</td>
<td>3.061</td>
<td>27.40**</td>
</tr>
<tr>
<td>Residual</td>
<td>8.737</td>
<td>6</td>
<td>1.456</td>
<td>13.12**</td>
</tr>
<tr>
<td>Stress x Trials</td>
<td>1.899</td>
<td>8</td>
<td>0.237</td>
<td>2.11*</td>
</tr>
<tr>
<td>Linear Components</td>
<td>1.957</td>
<td>1</td>
<td>1.957</td>
<td>9.52**</td>
</tr>
<tr>
<td>Quadratic Components</td>
<td>0.013</td>
<td>1</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>0.829</td>
<td>6</td>
<td>0.138</td>
<td>1.24</td>
</tr>
<tr>
<td>Anxiety x Trials</td>
<td>1.285</td>
<td>16</td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td>Linear Components</td>
<td>0.167</td>
<td>2</td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td>Quadratic Components</td>
<td>0.031</td>
<td>2</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>Stress x Anxiety x Trials</td>
<td>1.148</td>
<td>16</td>
<td>0.072</td>
<td></td>
</tr>
<tr>
<td>Linear Components</td>
<td>0.144</td>
<td>2</td>
<td>0.072</td>
<td></td>
</tr>
<tr>
<td>Quadratic Components</td>
<td>0.171</td>
<td>2</td>
<td>0.086</td>
<td></td>
</tr>
<tr>
<td>Groups x Trials</td>
<td>4.332</td>
<td>40</td>
<td>0.108</td>
<td></td>
</tr>
<tr>
<td>Error (b)</td>
<td>101.038</td>
<td>912</td>
<td>0.111</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>170.348</td>
<td>1079</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*P < 0.05

**P < 0.005
Fig. 1. Main effect of anxiety upon mean interjudge reliability. All means are in $z'$ values.
The interaction of stress and anxiety was not significant when considered independently of the trial variable. This indicates that if anxiety and stress do indeed influence the reliability of clinical judgments, the influence of these two variables is independent of each other.

There is a significant variation among the over-all trial means ($F=23.56$, $df=1 \& 912$, $p < 0.005$). The highly significant upward linear trend of the over-all trial means ($F=82.15$, $df=1 \& 912$, $p < 0.005$) indicates that learning took place across trials. The significant quadratic trend of the over-all trial means ($F=27.45$, $df=1 \& 912$, $p < 0.005$) indicates that the learning did not take place at a constant rate but is curvilinear. Figure 2 shows the graph of the overall trial means. The upward linear trend and the variation of rate of learning are both evident. The significant residual sum of squares ($F=13.12$, $df=6 \& 912$, $p < 0.005$) suggests that there may be trial position or stimulus set effects upon learning clinical judgment. The third hypothesis, that learning takes place, is supported.

The stress by trials interaction is significant ($F=2.14$, $df=8 \& 912$, $p < 0.005$) indicating that interjudge reliability is differentially influenced by the stress and neutral instructions over trials. Inspection of the graphical representation of this interaction in Figure 3 shows the stress Ss initially performing less well than the neutral Ss. The significant linear components ($F=9.54$, $df=1 \& 912$, $p < 0.005$) of the
Fig. 2. Mean interjudge reliability of all experimental subjects over trials.
Fig. 3. Interaction of stress and trials.
stress by trials interaction indicates a significant difference between the slopes of the two lines. Despite several crossings of stress and neutral curves the quadratic and residual components are not significantly different for the curves. These findings indicate that the curvature is not significantly different for the performance curves of the neutral and stress Ss. The stress by trials interaction and its linear components are consistent with a drive interpretation of the operation of stress. In hypothesis six only the initial decrement of performance of stress Ss was predicted.

Duncan's new multiple range test was used to test the significance of the differences between the means of all of the stress Ss and the means of all of the neutral Ss on a trial-by-trial basis. These means are shown in Table 4. The results indicated no significant difference in the performance of stress and neutral Ss on any single trial. While the results of the analysis of the trends of the stress by trials interaction supports a drive interpretation of the operation of stress, the results of the Duncan's tests suggest that the absolute differences between the stress and neutral groups on any single trial are not great. In interpreting the Duncan's range tests it must be borne in mind that they are based only on the means of a single trial. The trend analysis is based upon the aggregate of the means for all of the trials.

A graphical presentation of the anxiety by trials interaction is shown in Figure 4. While the anxiety by trials interaction was not ex-
Table 4

Mean Interjudge Reliabilities for all of the Stress and all of the Neutral Subjects over Nine Trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Stress</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.152</td>
<td>1.354</td>
</tr>
<tr>
<td>2</td>
<td>1.485</td>
<td>1.540</td>
</tr>
<tr>
<td>3</td>
<td>1.424</td>
<td>1.393</td>
</tr>
<tr>
<td>4</td>
<td>1.460</td>
<td>1.478</td>
</tr>
<tr>
<td>5</td>
<td>1.253</td>
<td>1.353</td>
</tr>
<tr>
<td>6</td>
<td>1.516</td>
<td>1.505</td>
</tr>
<tr>
<td>7</td>
<td>1.249</td>
<td>1.267</td>
</tr>
<tr>
<td>8</td>
<td>1.528</td>
<td>1.460</td>
</tr>
<tr>
<td>9</td>
<td>1.752</td>
<td>1.665</td>
</tr>
</tbody>
</table>
Fig. 4. Interaction of anxiety and trials.
pected to be significant as predicted in hypothesis four, the order of the means of the three anxiety levels is not as predicted. On the first three trials, MA Ss performed best and the HA Ss performed just slightly better than the LA Ss. The non-significant linear and quadratic components of the anxiety by trials interaction indicates that the crossover of the three curves for the anxiety levels is random. Hence, the effect of anxiety on the reliability of clinical judgment is not of significant proportions in this experiment.

The anxiety by stress by trials interaction was not significant indicating that the trial-by-trial means for the six experimental groups are not significantly different. The linear and quadratic components of the stress by anxiety by trials interaction are also non-significant. Furthermore, the non-significant groups by trials interaction indicates that the trial-by-trial learning curves of the six experimental groups are essentially similar.

Duncan's new multiple range tests were used to test the significance of the differences between the means of the six groups on a trial-by-trial basis. Only the means for the HA neutral and HA stress Ss on trial one were significantly different ($p=0.05$). The protection level against Type I errors for six means is 77 percent. Hence, this obtained first trial difference may be only a chance finding.

Figures 5 through 8 contain graphical presentations of a number of
High Anxious Neutral Subjects

Low Anxious Neutral Subjects

Fig. 5. Trial-by-trial means of the High Anxious and Low Anxious subjects tested under neutral instructions.
Fig. 6. Trial-by-trial means of the High Anxious and Low Anxious subjects tested under stress instructions.
Fig. 7. Trial-by-trial means of High Anxious Stress subjects and High Anxious Neutral subjects.
Fig. 8. Trial-by-trial means of Low Anxious Stress subjects and Low Anxious Neutral subjects.
pairings of the six performance curves that are the basis of the stress by anxiety by trials interaction. Separate figures are used only to facilitate visual assessment of the results.

Examination of Figure 5 shows that the performance of HA Ss tested under neutral conditions is superior to that of LA Ss tested under neutral conditions in six out of nine of the trials. The analysis of variance indicates that the crossovers of the two curves in Figure 4 are random and that the two performance curves are essentially similar. This finding partially supports hypothesis seven in that HA and LA Ss were not expected to perform differently under neutral instructions.

Figure 6 shows essentially identical performance curves for the HA and LA Ss who received stress instructions. Hypothesis seven called for a crossover of these two curves with the HA Ss initially performing less well than the LA Ss with the opposite relationship between HA and LA Ss expected by trial nine. The results do not support this portion of hypothesis seven. The consistent non-significant superiority of LA Ss performance over HA Ss performance on a trial-by-trial basis is remarkable in its uniformity and does agree with past research which shows that HA Ss find stress situations detrimental to performance when compared with LA Ss' performance (Walker, Neilsen, & Nicolay, 1965).

Except for a chance crossover between trials eight and nine, Figure 7 shows a slight non-significant trial-by-trial superiority of performance
of HA Ss tested under stress over HA Ss tested under neutral conditions. The relative non-significant trial-by-trial superiority (in eight out of nine trials) of LA Ss performing under stress compared to LA Ss who received neutral instructions suggests that LA Ss are better able to deal with stress (see Figure 8). The presence of an apparent large difference between trial-by-trial means of HA Ss tested under stress and neutral conditions (shown in Figure 7) relative to the small apparent differences indicated in Figure 8 for LA Ss tested under stress and neutral conditions suggests that HA Ss are relatively more susceptible to stress than LA Ss.

Table 5 contains a summary of the results of the stress Ss' pre-test ratings of how they felt about being shocked and their post-test responses to the question "How did you feel about being in a situation in which you might be shocked?" A sample of the pre-test rating scales and post-test Questionnaire are in Appendices B and C respectively.

Eighty-five per cent of all the stress Ss felt at least a little uneasy about being shocked. Included in the eighty-five per cent are Ss who reported feeling a little uneasy and those Ss who reported greater discomfort. The remaining fifteen percent of the Ss reported that they felt relatively calm about being shocked. No subjects reported feeling completely calm about being shocked. The mean ratings found in this study are about the same as those obtained by Schachter (1959). The results of the two ratings scales indicate that the Ss, on the whole, felt stressed. The reported pre-test level of felt stress is similar for
Table 5

Summary of Stress Subjects' Responses to the Pre-test Rating Scales and to the Post-test Question "How Did You Feel About Being in a Situation in Which You Might Receive a Shock?"

<table>
<thead>
<tr>
<th>Feeling About Shock</th>
<th>High Anxiety N=20</th>
<th>Middle Anxiety N=20</th>
<th>Low Anxiety N=20</th>
<th>Total N=60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Per Cent</td>
<td>No.</td>
<td>Per Cent</td>
</tr>
<tr>
<td>Felt Uneasy</td>
<td>18</td>
<td>90%</td>
<td>18</td>
<td>90%</td>
</tr>
<tr>
<td>Felt Relatively Calm</td>
<td>2</td>
<td>10%</td>
<td>2</td>
<td>10%</td>
</tr>
<tr>
<td>Mean</td>
<td>3.80</td>
<td></td>
<td>3.55</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.17</td>
<td></td>
<td>1.01</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Likelihood of Shock</th>
<th>High Anxiety N=20</th>
<th>Middle Anxiety N=20</th>
<th>Low Anxiety N=20</th>
<th>Total N=60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Per Cent</td>
<td>No.</td>
<td>Per Cent</td>
</tr>
<tr>
<td>Some Likelihood</td>
<td>17</td>
<td>85%</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td>Little Likelihood</td>
<td>3</td>
<td>15%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mean</td>
<td>3.35</td>
<td></td>
<td>3.55</td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.79</td>
<td></td>
<td>0.74</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post-test Question</th>
<th>High Anxiety N=20</th>
<th>Middle Anxiety N=20</th>
<th>Low Anxiety N=20</th>
<th>Total N=60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Per Cent</td>
<td>No.</td>
<td>Per Cent</td>
</tr>
<tr>
<td>Felt Stressed</td>
<td>18</td>
<td>90%</td>
<td>17</td>
<td>85%</td>
</tr>
<tr>
<td>Felt Unconcerned</td>
<td>2</td>
<td>10%</td>
<td>3</td>
<td>15%</td>
</tr>
</tbody>
</table>
the HA and MA Ss. The LA Ss showed somewhat less response to the stress instructions.

The stress instructions included statements indicating that only one in four Ss would receive electric shock. The second rating scale in Appendix B was designed as a pre-test measure of whether Ss felt they would be among those shocked. Only seven percent of all the stress felt that there was little likelihood of being shocked. No Ss felt that there was very little likelihood of being shocked. The remaining ninety-three percent felt there was at least some likelihood of being shocked. Included in the ninety-three percent are all Ss who felt that there was some likelihood of being shocked and all Ss who checked categories indicating greater likelihood of being shocked. The means and standard deviation of the stress Ss ratings suggest that LA Ss used less extreme ratings than the HA and MA stress Ss. The general indication is that prior to testing the stress Ss felt that they were among those who would receive shocks.

The responses to the open ended post-test question "How did you feel about being in a situation in which you might be shocked?" were quite varied. The responses were categorized on the basis of the Ss indicating that they had a stress reaction or felt calm. The results appear in Table 5. The results for all stress Ss combined are much like those for the pre-test rating scale. One LA S who reported feeling relatively
calm before the testing, in retrospect, reported feeling stressed. One MA S who reported feeling at least a little uneasy prior to testing, in retrospect, reported that he did not feel concerned as a result of being in a situation in which he might receive a shock. It seems that most stress Ss felt stressed both before and after testing.

The implications of particular responses to the open ended questions will be considered in the Discussion section. The one comment about the stress condition that seemed to stand out was the spontaneous remark by eleven out of sixty stress Ss that they felt stressed at the beginning of the series of clinical judgments but felt relaxed by the end of the series. All general psychology students were told at the beginning of the course that they could terminate participation in a particular experiment with no loss of experimental points if they felt the experiment was excessively stressful for them. Only one HA S dropped out of the experiment after hearing the stress instructions. He reported that he was very frightened. The E observed that this particular S was perspiring, his eyes were glassy, and his hands shook. This S was given reassurance and relaxed after a few minutes discussion with the E.
Chapter IV
Discussion

The outstanding finding in this study is the significant stress by trials interaction. The interjudge reliability of stress Ss was initially lower than that of neutral Ss but by the last trial the stress Ss' interjudge reliability was greater than that of the neutral Ss. The Duncan's multiple range test did not indicate a significant difference between stress Ss and neutral Ss on trial one or trial nine. The trend analysis of the stress by trials interaction is significant. Since the trend analysis is based upon the relative differences among all of the means in the stress by trials interaction, it is a more powerful test of the total set of means. The obtained effect of stress upon interjudge reliability was not hypothesized. It was expected (Hypothesis Six) that on initial trials stress Ss would perform less well than neutral Ss with the initial difference diminishing over trials.

It was hypothesized that under stress instructions HA Ss would initially perform less well than LA Ss with the relationship between the performance of HA and LA Ss reversing by the last trial. It was expected that the learning curves for the HA-stress Ss and LA-stress Ss would crossover. The anxiety factor had no significant effect upon interjudge reliability in this study. The obtained curves for the stress and
neutral Ss are like those predicted for the HA-stress and LA-stress Ss.

The stress manipulation in this study might be thought of as a behavioral manipulation of drive. The use of high and low anxiety Ss (as measured by the MAS) would be an experimental manipulation of drive by selecting two groups of Ss differing in organismic based drive. The results of the stress manipulation exactly fit Taylor's (1956) predictions for the relationship of manifest anxiety and complex task performance. In this study Taylor's theory is supported only for drive due to stress arousal.

Some of the previous studies relating clinical judgment and manifest anxiety have obtained only initial trial differences between HA and LA Ss but have not obtained the expected crossover of HA and LA performance curves. In these studies (Hunt & Blumberg, 1961; Hunt & Walker, 1963) the initial decrement of the HA Ss compared to LA Ss was attributed to HA Ss having more task irrelevant anxiety responses than the LA Ss. Hunt and Blumberg (1961) and Hunt and Walker (1963) found that HA and LA Ss' performance was much alike on later trials. Hence, their (Hunt & Blumberg, 1961; Hunt & Walker, 1963) results did not support the drive interpretation of anxiety (Taylor, 1956) but did support the drive stimulus interpretation of anxiety (Child, 1954).

In the earlier studies which did support the drive stimulus interpretation of anxiety (Hunt & Blumberg, 1961; Hunt & Walker, 1963) the Ss were tested individually. It was hypothesized that the amount of task
irrelevant anxiety responses were a positive function of the amount of stress in the testing situation (Spence, 1963). On the assumption that a group testing condition was less stress arousing than an individual testing condition, Walker, Hunt, and Schwartz (1965) hypothesized that HA and LA Ss would perform a clinical judgment task equally well if they were tested in groups. Their hypothesis was supported.

Pribyl, Walker, and Hunt (1965) and M. K. Pribyl (1966) had HA and LA Ss perform clinical judgment tasks in an individual situation but did not find the expected initial trial difference between the HA and LA Ss. Pribyl et al. (1965) also tested HA and LA Ss in a group condition and found no difference as expected. Perhaps the inconsistency results of studies in which HA and LA Ss performed clinical judgment tasks individually is due to only weak or transitory stress being produced in the individual testing situation.

The strong stress manipulation used in this experiment was effective according to subject's reports both before and after the measurement of the dependent variable. In addition, the stress instructions had a significant effect upon interjudge reliability over trials. Taylor's (1956) and Hunt and Walker's (1963) drive theory interpretation of manifest anxiety did not receive support in that the expected crossover of the performance curves of the HA and LA Ss tested under stress conditions was not obtained (see Figure 6). The results of this study do not support the drive stimulus interpretation of anxiety in the neutral con-
dition. That is, there is no significant initial trial difference be-
tween HA and LA Ss tested in the neutral condition (see Figure 5).

While the results of this study lend support to the drive inter-
pretation of anxiety (Taylor, 1956), they do so only when drive is aroused
through stress.

The problem of the drive stimulus versus a pure drive interpretation
of anxiety is as yet unsettled (Hunt & Walker, 1965; Sarason, 1960; Spence,
1963; Taylor, 1956). It is quite possible that the clinical judgment
studies which have supported the drive stimulus interpretation of mani-
fest anxiety (Hunt & Blumberg, 1961; Hunt & Walker, 1963; Walker, Hunt,
& Schwartz, 1965) have tapped only a portion of the learning curves ex-
pected in the drive theory interpretation of manifest anxiety. The
previously obtained initial trial difference between HA and LA Ss is
predicted by the drive interpretation. The empirical learning curves
of the HA and LA Ss have then come together. Possibly, the reason the
HA Ss' performance does not surpass that of the LA Ss' performance is
that learning becomes asymptotic. Perhaps even ten trials is not suf-
ficient number of trials to permit crossover to take place. M. K. Pribyl
(1966) has assembled another set of stimuli which might be used to assess
the effect of manifest anxiety upon clinical judgment over a greater
number of trials.

While the interaction obtained in this study between stress and
trials does correspond with drive theory, there is another explanation
which fits the obtained results quite well. Several of the stress Ss (11 out of 60) spontaneously stated that they felt stressed at the beginning of the clinical judgment task but found that they became more relaxed as they progressed through the task. Anxiety may have interfered with performance of stress Ss early in the task but once they relaxed their relatively greater involvement in the task increased their efficiency relative to the neutral Ss. Ego involving instructions have had a positive effect upon performance in some studies (Sarason, 1960). Several stress Ss commented that they felt stressed but also felt that they were motivated to take more care to accurately perform the experimental task. At this time efforts should be made to exactly replicate this study to determine the generality of the significant stress by trials interaction obtained in this study. Should the present findings be obtained in a replication, carefully designed stress and ego-involving instructions should be constructed and used along with pre-test and post-test rating scales and questionnaires in a study pitting the effects of ego-involvement on clinical judgment against the effects of stress on clinical judgment.

The over-all performance of the MA Ss was nonsignificantly higher than the Ss of either of the other two anxiety levels (see Figure 1). No specific theory was used in hypothesizing that the performance of the MA Ss would be intermediate to that of the HA and LA Ss. The MA group
was tested only to explore the performance of Ss who score in the middle ranges of the MAS scale. Malmo (1962) equated manifest anxiety as measured by the MAS and "activation." Activation as Malmo used the term refers to an intensive dimension of arousal. Another conception of activation used by Malmo is that it refers to a generalized drive state which does not have any "directing" properties. Malmo reviews many experiments which support the theory that increasing Ss activation from a low level to a high level causes the Ss performance to initial increase to an optimum level and then decrease. The performance curve as a function of increasing activation would be in the form of an inverted U. While the relationship of the over-all performance of the three anxiety groups is not a statistically significant one, the form of the graphical presentation (Figure 1) does fit Malmo's conception of activation.

Malmo (1962) suggests that activation phenomena would be most obvious in intrasubject comparisons of performance on the same task under differing levels of activation. Malmo's suggestion is based on research showing that intersubject variation of activation level and performance for a given situation and task are far greater than intrasubject variation of activation level and performance for a given situation and task over time. The nonsignificant differences in performance among the Ss of the three anxiety levels may be due to the high intersubject variation.

The variation that exists in anxiety scale scores may reduce the reliability of assigning Ss to a particular segment (for example, middle
anxious) of a scale. Perhaps, this is one source of intersubject variation reducing the empirically obtained effects of activation upon performance. Cronbach (1960) has demonstrated that the use of several tests with a "selective sieve" technique may substantially increase the reliability of designating persons as belonging in a particular category. The several anxiety scales described by Sarason (1960) might profitably be used with a "selective sieve" technique to identify "high", "middle," and "low" anxiety Ss in future studies. Future research relating clinical judgment and manifest anxiety might profitably use a co-variance analysis of the accuracy of judgments and the subject's MAS score. Another variable that might be investigated is the co-variation of the subject's intelligence and the accuracy of his ratings of confusion.

The clear learning effect across trials which was obtained only in this study and in the only other previous clinical judgment study using automated constant time interval presentations of stimuli (Mary K. Pribyl, 1966) suggests that the stimulus presentation time interval may be a critical variable. Future investigations might profitably manipulate this variable and provide a link between clinical judgment and the vast literature concerned with the effects of massed and distributed practice upon learning (Underwood & Richardson, 1957; Underwood & Schulz, 1961).

In this study each S's rating was correlated with the mean ratings for his group on a trial-by-trial basis. If the standard deviations of the Ss' ratings were high on the initial trial and went down across trials the increase correlations across trials might be the spurious re-
sult of regression toward the mean. The above case would be established if Ss made relatively more extreme ratings on early trials than on later trials. That this is not the case is suggested by previous studies in clinical judgment (Blumberg, 1961; Hunt & Blumberg, 1961; Pribyl, Mary K., 1967) which indicate that the mean interjudge reliabilities agree well with validity measures. That is, the Ss ratings do not approach the mean as they progress across trials, but become more like the ratings given by the experienced clinicians. Future research should directly refute the regression hypothesis by determining whether the mean ratings, the standard deviations of the ratings, and the use of judgment categories varies across trials in such a manner that a spurious increase in correlation is obtained.

While learning across trials seems to be a general phenomenon, each study using the ratings of confusion exhibited in schizophrenics' vocabulary test responses has revealed considerable variability in judge's ratings due to many unknown parameters of the sets of stimuli. In this study a significant residual variance was obtained in the trials effect after the variance due to linear and quadratic trends was subtracted from the variance attributable to trials. The significant residual variance may be due to differences in the sets of stimuli used in each trial. Mary K. Pribyl (1966, 1967) has demonstrated that the standard deviation of the stimuli (that is, the standard deviation of the clinician's ratings used to standardize the stimuli) has a significant effect
upon interjudge reliability. Studies designed to identify and scale the
relevant parameters of the currently extant stimuli are indicated. The
extensive work carried out in verbal learning by Underwood and his co-
workers would be a relevant source of design and experimental techniques
(Underwood & Schdlz, 1960).

Lately, there has been concern about the effects of experimental
stress and deception upon subjects (Baumrind, 1964; Kelman, 1967). The
main purpose of giving Ss who participated in this study an opportunity
to discuss the experiment after all the data were collected was to per-
mit identification of Ss who were adversely affected by the stress in-
structions. If any Ss had been adversely affected, appropriate amelio-
ratve actions would have been taken. The fact that no Ss took advan-
tage of the opportunity to obtain feedback concerning the experiment
suggests that no one was adversely affected by the stress instructions.

As the amount of literature taking a single approach to a research
problem grows, the number of Es increase. The particular approach to
clinical judgment used in this study has not been an exception. The Es
have ranged in educational background from advanced undergraduate psychology majors, through Ph.D. candidates, to full-time faculty members.
Experimenter-related variables have been shown to have an effect on the
results of experiments in other areas (Rosenthal, 1964). Future invest-
tigations in clinical judgment might well assess the relationship of E
variables to clinical judgment.
Chapter V

Summary

This study investigated the influence of manifest anxiety and stress upon clinical judgments. Previous research in this area has not supported the drive theory interpretation of manifest anxiety. On a complex task (for example, clinical judgment) Taylor (1956) predicts that high anxious (HA) Ss will initially perform less well than low anxious (LA) Ss with the performance difference disappearing with repeated trials and the HA Ss eventually performing better than the LA Ss. Assuming that drive level was not sufficiently high in previous studies, stress instructions were used to heighten drive of Ss of three levels of anxiety while performing judgments. Support of the drive interpretation of anxiety was expected with the heightened drive consequent to stress.

Male undergraduate Ss (N=120) rated the amount of confusion exhibited in ninety vocabulary test responses from schizophrenics' test protocols. One-half the Ss in each anxiety level (40 HA, 40 middle anxious, and 40 LA as defined by scores on the Taylor MAS) were threatened with electric shock (no shocks were given) prior to task instructions. An analysis of variance was done on the z' values of the mean interjudge reliabilities (a measure of interjudge agreement), the variables being stress, anxiety level, and trials (nine trials of ten stimuli). Trend
analysis was done on the trial-by-trial means.

Stress had an effect on clinical judgment with the performance curve of the stress Ss lower than that of neutral Ss on initial trials. Stress Ss' performance surpassed the neutral Ss' performance on later trials. Manifest anxiety had no significant effect upon clinical judgment. There was a significant upward linear trend for the over-all trial means, indicating that learning took place across trials. The drive interpretation of anxiety was supported only for drive induced by stress.
References


Hunt, W. A., & Volkmann, J. The anchoring of an affective scale. 

Hunt, W. A., & Walker, R. E. Manifest anxiety and clinical judgment--

Hunt, W. A. & Walker, R. E. Factors influencing clinical judgment. 
Final report on Office of Naval Research contract Nonr-1228 (18) 
with Northwestern University, August, 1965.

Johnson, D. M. The psychology of thought and judgment. New York: 

Jones, N. F. Context effects in judgment as a function of experience. 

Kelman, H. C. Human use of human subjects: The problem of deception in 

Malmo, R. B. Activation. In Bachrach, A., (Ed.) Experimental foundations 
386-422.


Meehl, P. E. Clinical vs. statistical prediction. Minneapolis: Univ. 

Pribyl, J. F. Judgment as a function of manifest anxiety and social 

Pribyl, J. F., Walker, R. E., & Hunt, W. A. Judgment as a function of 
manifest anxiety and social conditions. Percept. mot. Skills, 1965, 
21, 759-765.

Pribyl, Mary K. The effect of anxiety and feedback on clinical judgment. 

Pribyl, Mary K. The effect of learning on reliability and validity of 
University, 1967.

Rabbie, J. M. Differential preference for companionship under threat. 


Equipment Used in Stress Manipulation
Appendix B

How do you feel about being shocked?

Put a check mark on the scale below to indicate your feelings about being shocked.

I feel extremely uneasy
I feel very uneasy
I feel quite uneasy
I feel little uneasy
I feel relative- complete- uneasy
I feel ly calm
I feel ly calm

(6)* (5) (4) (3) (2) (1)

Put a check mark on the scale below to indicate how likely it is that you will receive electric shocks.

Very great likelihood
Great likelihood
Some likelihood
Little likelihood
Very little likelihood

(5)* (4) (3) (2) (1)

*The scale values assigned to the various categories were not shown on the rating scales used by the experimental subjects.
Appendix C

Questionnaire

1. Did you know anything about this experiment beforehand? Yes ___ No ___

2. If the answer to the above question was yes, the subject was asked: "What did you know about it?" ____________________________

3. Did you understand what you were expected to do? Yes ___ No ___

4. If the answer to the previous question was no, the subject was asked: "What didn't you understand?" ____________________________

5. Do you have any questions or comments about the experiment? ____________________________

6. If the stress subjects did not spontaneously comment on how they felt about being in a situation in which they might be shocked, they were asked: "How did you feel about being in a situation in which you might receive a shock?" ____________________________
The dissertation submitted by Joseph F. Pribyl has been read and approved by members of the Department of Psychology. The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

[Signature of Adviser]

[Date]