METHODOLOGICAL STUDIES IN DETECTION OF DECEPTION

Richard I. Thackray and Martin T. Orne

Unit for Experimental Psychiatry

Institute of the Pennsylvania Hospital

Philadelphia, Pennsylvania

This investigation was supported by

U. S. Army Medical Research and Development Command,

Department of the Army, under Research Contract No. DA-49-193-MD-2647

SUMMARY

Two studies were conducted employing both innocent and guilty Ss. The guilty condition was established by having Ss play the role of an enemy espionage agent (courier). Innocent and guilty Ss were separately tested in groups of six or seven Ss. Both relevant-irrelevant (RI) and peak of tension (PT) methods were employed. GSR response to each question was separately recorded for each S. In addition, a composite GSR was also derived from all the Ss in each group.

In the first study, separate subgroups were given differential feedback of information regarding the polygraph's effectiveness during a preliminary phase of the interrogation. Results indicated that those Ss receiving information that the polygraph was effective in detecting their lies were detected more readily during the subsequent actual interrogation than were those Ss receiving information that the polygraph was ineffective. This effect was true only for the PT series, however. For the RI method of presentation, differential feedback of information did not influence detection rate. When the overall distributions of guilty and innocent Ss were compared, it was found that the courier paradigm was capable of yielding 83 percent correct classification of innocent and guilty Ss. Plotting each innocent and guilty group on the basis of their group GSR's, however, resulted in virtually no overlap between innocent and guilty groups.

In the second study, no differential treatment of innocent and guilty Ss was included in the design in order to provide a clearer interpretation of the effectiveness of the courier paradigm itself. The design was also modified to allow a more adequate comparison of RI and PT methods of interrogation. In addition, information was sought relative to the question of whether a S can be detected when he is informed that his responses are not being evaluated.

Results of the second study confirmed the findings of the previous experiment relative to the effectiveness of both the courier paradigm and the group GSR. Using the group GSR, the distributions of innocent and guilty groups were clearly separate with no overlap.

The second experiment also demonstrated a significant difference between RI and PT methods in favor of the RI. This was shown to be generally in accord with previous findings of this laboratory.

Regarding detection efficiency when S is "unaware," the results were inconclusive. However, the obtained data were in the predicted direction and suggest the need for further research.

EXPERIMENT I

Introduction

The detection ratios achieved in different studies have varied widely (Ellson et al., 1952; Kubis, 1962; Lykken, 1959; Lykken, 1960). Previous research conducted by this laboratory has demonstrated that motivation, role perception, and method of presentation account for much of this variance (Gustafson & Orne, 1963; Gustafson & Orne, 1964; Gustafson & Orne, 1965). However, it was felt essential, for future investigations on the determinants of detection, to develop a laboratory paradigm which would yield high, extremely stable detection ratios. Because of the importance of motivational factors, it was desired to develop a paradigm which maximized the degree of involvement which S had with the experiment. Further, it appeared desirable to develop a paradigm which would allow the E to assume the traditional role of interrogator. This would demand Ss be instructed by another E previously. Such a procedure would raise the level of motivation of Ss on the one hand and, on the other, would facilitate the use of techniques during the interrogation procedure which might maximize responsivity.

The present experiment was designed to utilize two Es in order to maximize involvement, permit an evaluation of a group GSR technique, and finally to specifically explore the effect of a harmless but painful electric shock, or the threat of shock, on the detection of deception.

The approach was to run separate innocent and guilty groups of six or seven Ss each. The guilty groups were taught information in an ego-involving fashion by one E, then turned over to another E for interrogation. Innocent groups, of course, were taught none of this information. Prior to the actual interrogation, Ss selected a card and were told that when questioned about the card, the polygraph would automatically deliver a shock whenever it detected any evidence of lying. Two of the Ss in each group were shocked for the right card, two for the wrong card, and two received no shock. This design was intended to separate the effect of shock itself on subsequent detection from the cognitive aspects of feedback from the polygraph. Thus, two subgroups would receive shock, one of which would be told in effect that the machine correctly picks up lying while the other group would be "informed" that the machine does not work accurately. The shocked-incorrectly group and the no-shock group would both be informed that the machine was working inaccurately but, of course, would interpret this information differently. It was predicted that the subgroup receiving feedback that the polygraph was effective would be easier to detect in the subsequent interrogation than the subgroup led to believe the polygraph was not detecting anything. Further, it was predicted that those Ss receiving conflicting feedback (the shocked-incorrectly subgroup) would occupy an intermediate position in terms of their subsequent detection.

METHOD

Subjects

Twenty groups containing 7 Ss each were scheduled from local universities. Of the 20 groups, 10 were randomly assigned to the "courier" or guilty condition and the other 10 to the "innocent suspect" condition. Although 7 Ss were scheduled for each group, some of the groups tested contained only 6 Ss because of failures to report for the experiment. Consequently, the 10 courier groups contained a total of 66 Ss while the 10 innocent groups contained 68 Ss.

Procedure

Upon reporting to the lab, a given group was taken to a room and played a tape-recording appropriate to their group by E_1. Those assigned as couriers were played a recording telling them that they were to play the role of an enemy courier who knows six code words. The code words form the key to deciphering a message carried by a separate courier, and without the code words, it would be virtually impossible to decipher the message. They were told that they would be interrogated later by another person (E_2) who would attempt to determine whether or not they actually were couriers. The Ss were then given booklets which were constructed in such a way as to maximize over-learning of the code words. Ss who did not have perfect retention of all code words at the completion of the hour learning session were not used in the remainder of the experiment. Only one S, however, had to be discarded for this reason.

The Ss in the innocent suspect groups were played a tape explaining that they were under suspicion of being enemy couriers and would be interrogated in the same manner as their counterparts in the real courier groups. Instead of then being given booklets containing the code words, however, they spent an hour completing a short-answer type questionnaire containing items relating to attitudes toward various aspects of lie detection. (The questionnaire data will be reported in a separate paper.)

At the completion of this first hour, another tape-recording was played. If the Ss constituted an innocent group, the tape informed them that they would now be interrogated, but since they had nothing to hide, they had no reason to fear the procedure. For a guilty group, the recording emphasized that the interrogator would do his best to obtain a confession, and that it would be difficult to deceive him. However, it was pointed out that it was possible to beat the polygraph, and that emotionally mature, intelligent individuals were able to succeed.

For the actual instructions and directions contained in the tape-recordings, the reader is referred to the Appendix.

The particular group was then turned over to E_2 (the interrogator) who had no prior knowledge as to whether it was an innocent or a courier group. E_2, however, did know the six code words.

After Ss were randomly assigned to chairs in the interrogation room, E_2 went to an adjacent room, and a tape was played which

attempted to convince the group that if they actually were couriers they might as well confess now because the polygraph examination would reveal their guilt or innocence with absolute certainty. They were then told that each S would shortly be asked to draw a card for the purpose of demonstrating the polygraph's effectiveness. The group was informed that when a person lied concerning a card which he had drawn, the polygraph was so constructed that it would automatically deliver an electric shock to that person.

At the completion of this tape, E_2 re-entered the Ss' room and each S was given the opportunity to confess to being a courier. (No S ever confessed to being a courier either at this time or at any other time during the experiment proper.) S then drew a card from a deck of playing cards and memorized the suit and face value. E_2 left the room and each S wrote the value and suit of the card he had drawn on a slip of paper and placed it below his chair. After Ss were finished, E_2 re-entered the room along with an assistant. The assistant attached GSR and shock electrodes to each S along with blindfolds, headphones, and dummy microphones.[2]

While the assistant was thus occupied with the Ss, E_2 left the room and assigned Ss to the various shock conditions. Two Ss were assigned to a "shocked for correct card" (SCC) condition, in which shock was administered after the suit or face value of a card which they actually drew was mentioned; another two Ss were assigned to a "shocked for incorrect card" (SICC) condition, in which shock was

administered after the suit or face value of a card which they had not drawn was mentioned; and a final two or three Ss (depending on whether the total group contained six or seven Ss) were assigned to a "no shock" (NS) condition, in which no shock was administered for any card. The deck from which the Ss drew their cards was constructed in such a way that E_2 knew the cards selected by two of the Ss. These two Ss were assigned to the SCC condition. Assignment to the other conditions was completely random.

After all electrodes and the rest of the equipment were attached, each S was individually administered the electric shock. This was done both to demonstrate the nature of the shock which they could expect to receive when they lied, and also to insure that each S was properly connected to the shock apparatus. All Ss received the same physical intensity of shock. The level chosen was definitely uncomfortable, but not painful.

A tape-recording was then played in which the Ss were questioned concerning the cards which they had drawn. A relevant-irrelevant (RI) method of presentation was employed in which each of the four possible suits and each of the thirteen possible face values of the cards was mentioned once. Presentation was random with the interword interval varying randomly between eight to fifteen seconds. Ss were informed that they had to answer "no" to each question, and, therefore, could expect to receive a shock when questioned concerning their particular card. Shocks were administered to the appropriate individuals

immediately following their response.

After completion of the questions concerning the cards, E_2 entered the Ss' room and gave each S a final opportunity to confess to being a courier. As was previously stated, no S ever confessed to being a courier during the experiment proper. Following this final opportunity to confess, E_2 left the Ss' room.

The interrogation tape-recording was played in which Ss were told that they would now be questioned concerning certain words. They were told to answer truthfully in order to avoid further possible shocks. No shocks, however, were administered at any time during the remainder of the interrogation.

After an initial "buffer" question was asked, the interrogation was begun with the question "Were any of the following words critical code words?" The RI method, used during the first phase of the interrogation, included the five critical words, three "confusing" words (these were words which appeared in the original training booklet, but were stated to be unimportant to the interrogation procedure), and six neutral words. All words were administered in randomized order. Interword intervals varied randomly from eight to fifteen seconds.

Following the RI phase, Ss were told that they would be questioned about various colors. A modified peak of tension (PT) method was employed in which the color "green" (one of the code words) appeared as the fifth color in a series of six.

At the completion of the experiment, and while the electrodes and rest of the equipment were being removed, Ss were told that the experiment was completely finished and that they could now tell E_2 whether they were guilty or actually innocent Ss. However, as testimony to the degree of involvement of these Ss, very few were willing to admit to being either innocent or guilty even when every attempt was made to convince them that the experiment was really completed.

Apparatus

An Offner Type R Dynograph was used to record GSR responses. Two silver EEG electrodes were taped to the index and middle fingers of each S's hand with the leads going to individual bridge circuits. The output of each bridge was AC coupled to one of the channels of the Offner. A constant gain setting was used for all Ss.

To obtain the group GSR, each preamplifier output from the individual GSR channels was coupled through separate 1 megohm resistors to the input of the driver amplifier on the eighth channel. Thus, the voltage appearing across the input of the eighth or group GSR channel was the voltage produced by the sum of the currents flowing through each of the 1 megohm resistors in the individual preamplifier outputs.

The shock apparatus consisted of a simple inductorium with the output going to seven switches to allow administration of the shock

to each S individually. Electrodes consisted of two, two-centimeter copper disks attached to the ventral surface of the forearm of each S. A constant intensity was used for all Ss with intensity set at a level sufficient to produce a mild degree of tetany. Shock duration was approximately .5 seconds.

RESULTS

The basic response data of this study were the changes in skin resistance (GSR's) produced by the verbal stimuli. Magnitude of GSR was obtained by measuring in millimeters the maximum pen deflection to occur within three seconds following the onset of each verbal stimulus.

The initial technique employed for analyzing the data consisted of computing each S's mean GSR response to both critical words and neutral words, and the ratio of critical to neutral response then obtained. This procedure was followed for the RI series, while for the PT series, each S's GSR response to the color green (the critical word) was divided by his mean response to the other five neutral colors.

Subgroup Effects

Figure 1 shows the ratios obtained for all Ss in each subgroup in the RI series.[3] Only the courier subgroups are displayed in this Figure as well as in Figures 2-5, since only with these Ss would the differential subgroup treatments be expected to influence the ratios. No effect of the treatments on the subgroup means is apparent in

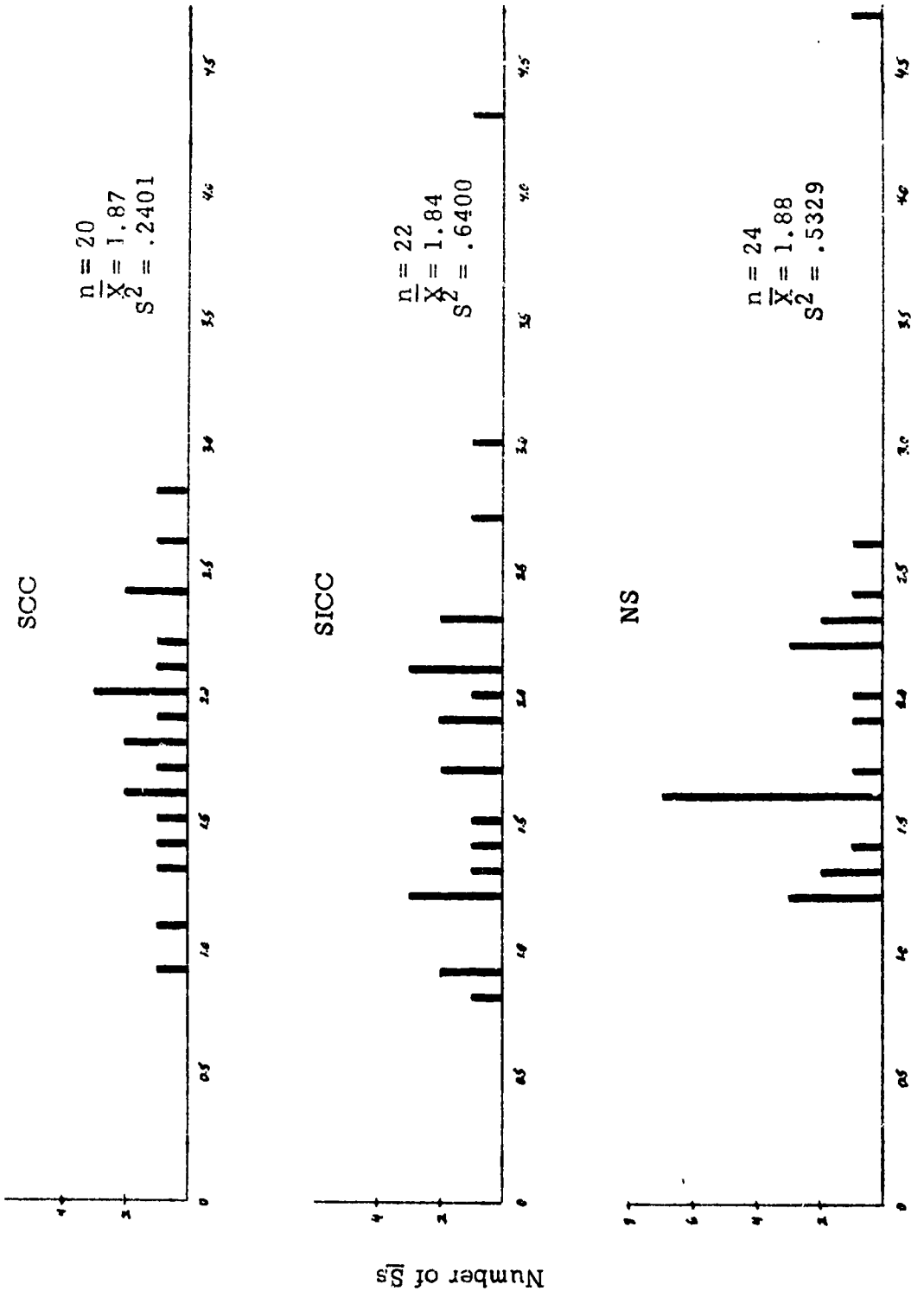

Figure 1. GSR ratio of critical to neutral words for all Ss in each subgroup in the RI series.

this Figure, and a Kruskal-Wallis Test applied to the data yielded an H value of .332 ($p > .05$).

Figure 2 displays the ratios for all Ss in each subgroup in the PT series. A Kruskal-Wallis Test applied to these means revealed the differences to be significant with an H value of 9.41 ($p < 0.1$).

In order to determine whether combining the RI series with the PT series might result in improved subgroup differentiation, both series were converted to z score distributions and each S's z score averaged. However, a Kruskal-Wallis Test applied to the averaged z scores yielded a value ($H = 3.72$, $p > .05$) which was substantially less than that previously obtained for the peak of tension series. These data are shown in Figure 3.

It is of some interest to note that in Figures 1, 2, and 3 the SICC subgroup consistently shows the greatest variance while the SCC distribution displays the smallest. F tests applied to the variances of these two subgroups yielded values of 2.67, 2.44, and 3.69 for the RI, PT, and combined series respectively, all of which exceed the .05 level of significance. In addition, in both PT and combined series, the variances of the SICC subgroup are significantly larger than the variances for the NS subgroup ($F = 2.44$, $p < .05$; $F = 2.56$, $p < .05$).

It is apparent from these data, then, that only in the PT series does differential treatment of the three subgroups have any effect upon the means. In this series, shock significantly increases detection over the NS condition, but examination of Figures 2 and 3 clearly reveal that

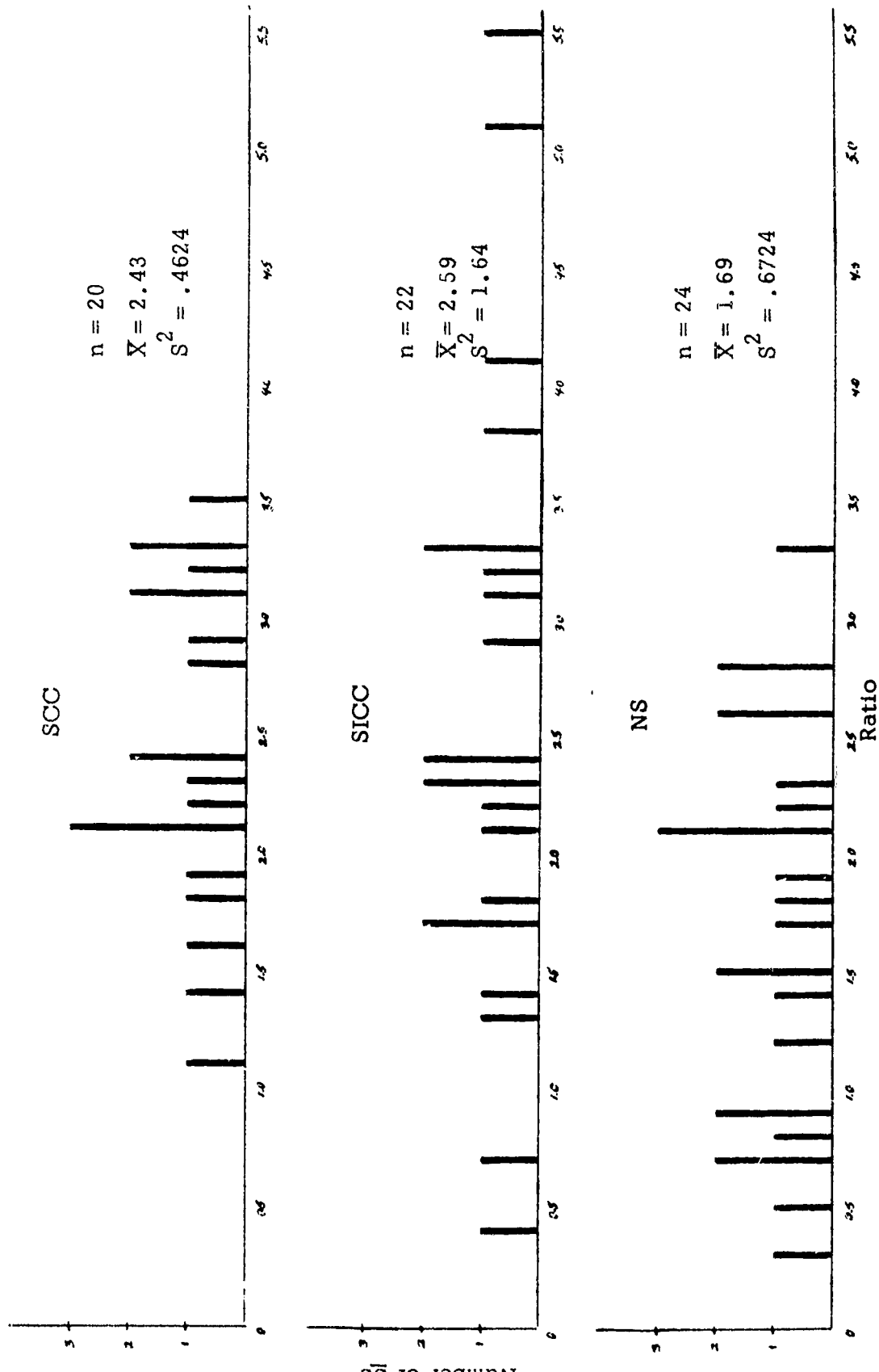

Figure 2. GSR ratio of the critical word to the mean of the neutral words for all Ss in each subgroup in the PT series.

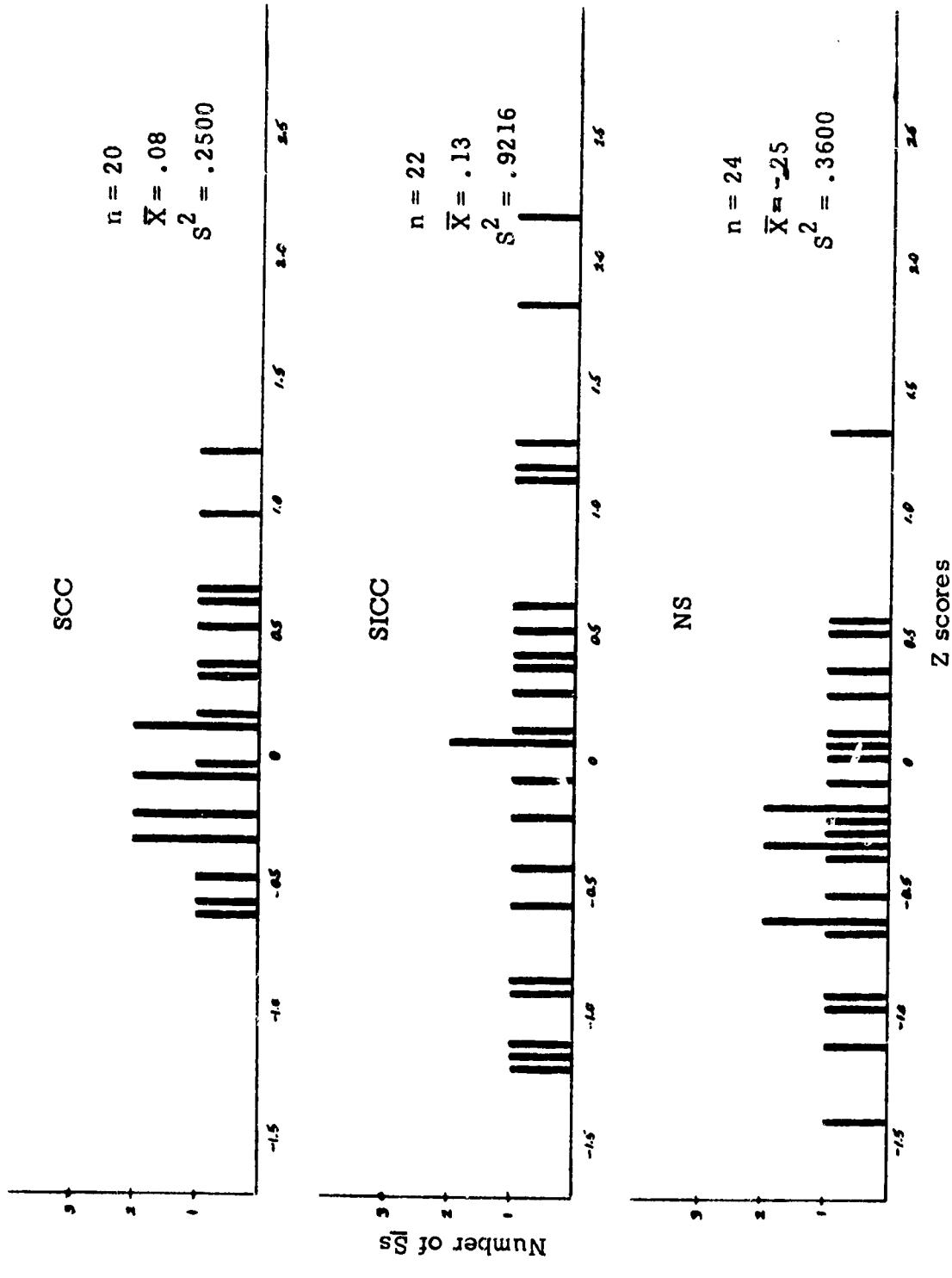

Figure 3. Mean Z scores for Ss in each subgroup in the combined RI and PT series.

the two shocked groups do not differ in mean response among themselves. The inflated variance of the SICC relative to the SCC subgroup is the only indication of possible differential behavior among these two subgroups.

Although it was originally intended to utilize only ratio scores in analyzing the results, it was decided to explore the use of ranking methods as an alternate technique. Consequently, in the RI series, the GSR for each of the five critical words was separately ranked against the six neutral words, with a rank of 1 assigned to the largest response. The five ranks thus obtained were averaged for each subject and the results shown in Figure 4. Although the means of the three subgroups are in the predicted direction, a Chi square Test performed on the data was not significant (see Table 1). Application of a ranking method to the PT series, however, did yield a significant Chi square (see Table 2). These latter data, which are displayed in Figure 5, more clearly reflect the predicted trends, with the percent of Ss in each subgroup receiving a maximum rank being 95, 73, and 50 percent for the SCC, SICC, and NS subgroups respectively. Separate comparisons were then made between each subgroup to determine the significant effects. The comparison of SICC with NS resulted in a Chi square of 1.63 ($p = .20$). Comparisons of SCC with NS, and SCC with SICC yielded Fisher Exact Probability values of .002 and .056 respectively. Thus, it is evident that a significant difference exists between the SCC and NS subgroups. No difference exists between the NS and SICC subgroups although the obtained probability value of .056 for the SICC and SCC comparison reaches borderline significance.

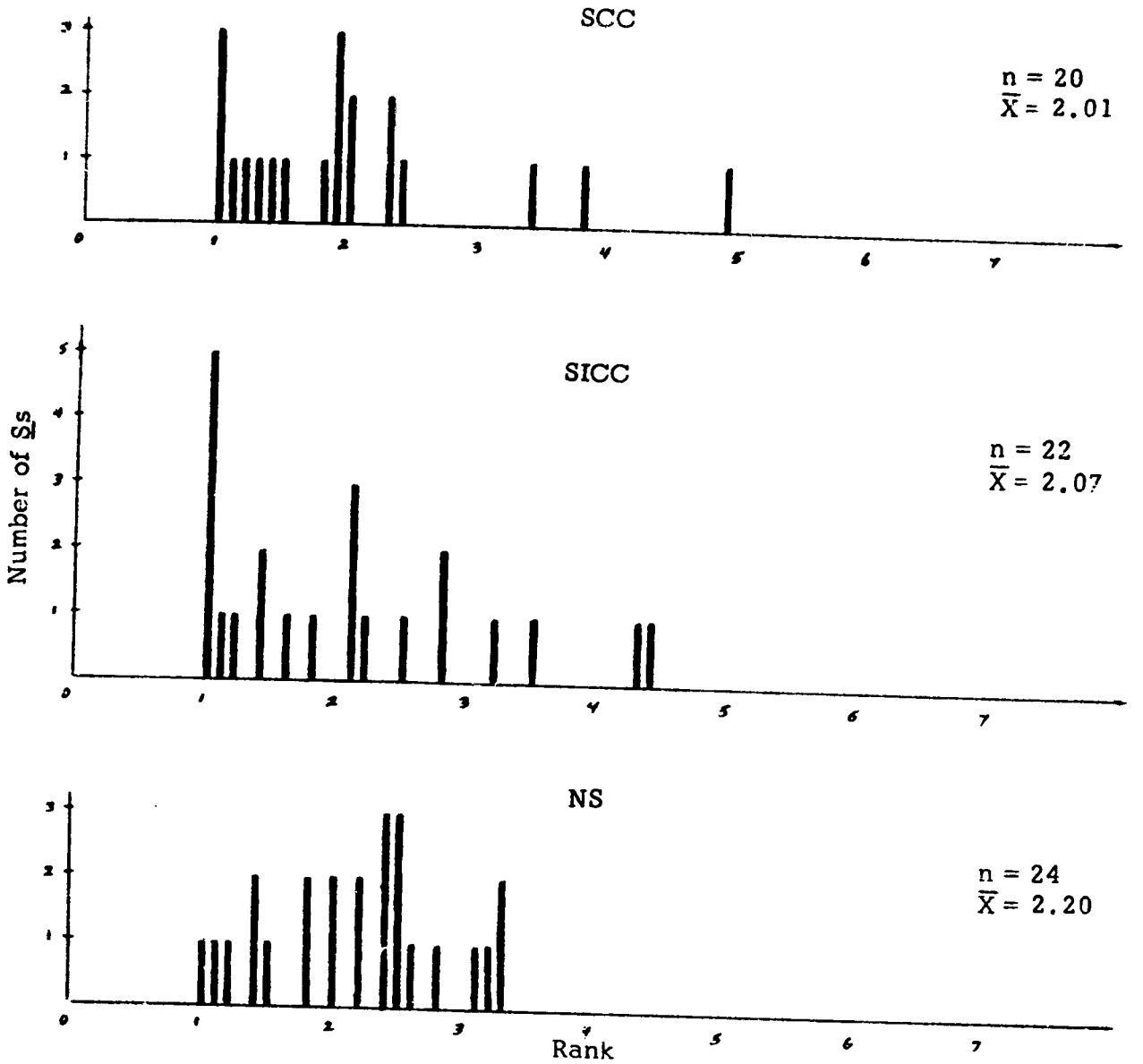

Figure 4. Mean GSR rank of the five critical words relative to the six neutral words in the RI series.

Table 1

Number of Ss in each subgroup in the RI series having a mean GSR rank falling above or below rank 2.

Ranking category	Subgroup			Chi square
	SCC	SICC	NS	
2 and below	14	11	10	3.63
Above 2	6	11	14	$p > .10$

Table 2

Number of Ss in each subgroup in the PT series whose GSR rank on the critical word falls at or below rank 1.

Ranking category	Subgroup			Chi square
	SCC	SICC	NS	
1	19	16	12	10.87
Above 1	1	6	12	$p < .01$

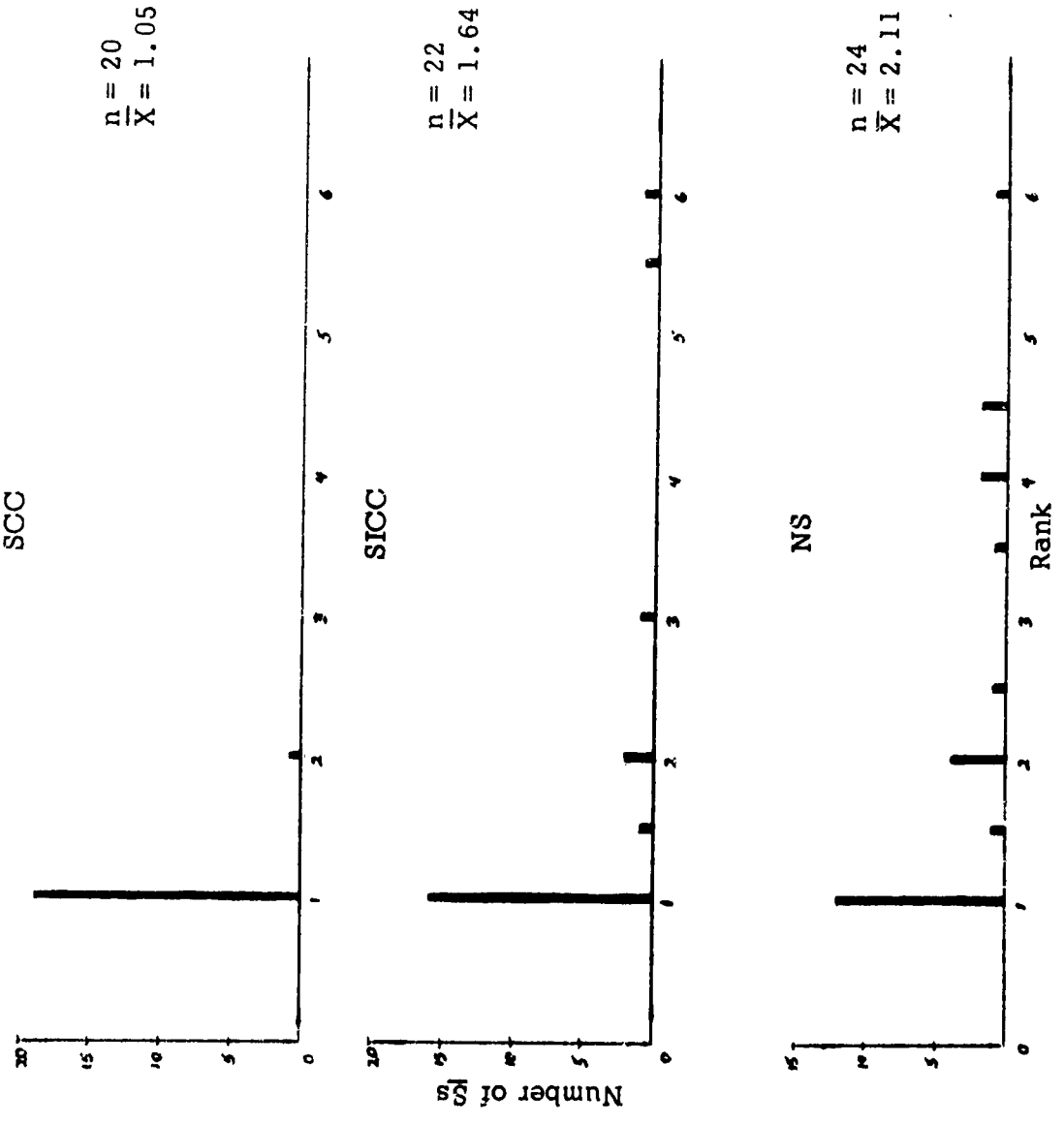

Figure 5. GSR rank of the color green (critical word) relative to the other colors in the PT series.

Overall Distributions of Guilty and Innocent Ss

Figure 6 presents the GSR ratio of critical to neutral words for all guilty and innocent Ss in the RI series. The ratio distributions for the PT series are displayed in Figure 7. Although the mean for the guilty Ss clearly exceeds the mean for innocent Ss in both series, considerable overlap between the distributions is shown in both Figures.

The RI data plotted according to ranks are shown in Figure 8. In order to allow comparison of the relative efficiency of the two methods of scoring the data, and also to obtain some index of the effectiveness of the courier paradigm, the point was determined for both the low end of the guilty distribution and the high end of the innocent distribution which yielded an approximate equal percentage of overlap. The number of Ss in each of these two tails was subtracted from the total number of Ss in both distributions. The ratio of Ss "detected" to the total number of Ss was then computed and the resulting number expressed as a percentage. For the data displayed in Figure 8, the detection ratio was found to be 84 percent. This same percentage was also obtained for the ratio scores shown in Figure 6. In Figure 9, which displays rank scores for the PT series, a detection ratio of 83 percent was obtained, while the PT data scored according to ratio scores (Figure 7) yielded a detection ratio of 82 percent. It is thus apparent that both methods of scoring the RI and PT data are equally effective, at least for these distributions.

Group GSR

Figures 10 and 11 show the group GSR ratios for the RI and PT series respectively. The guilty and innocent distributions in both series

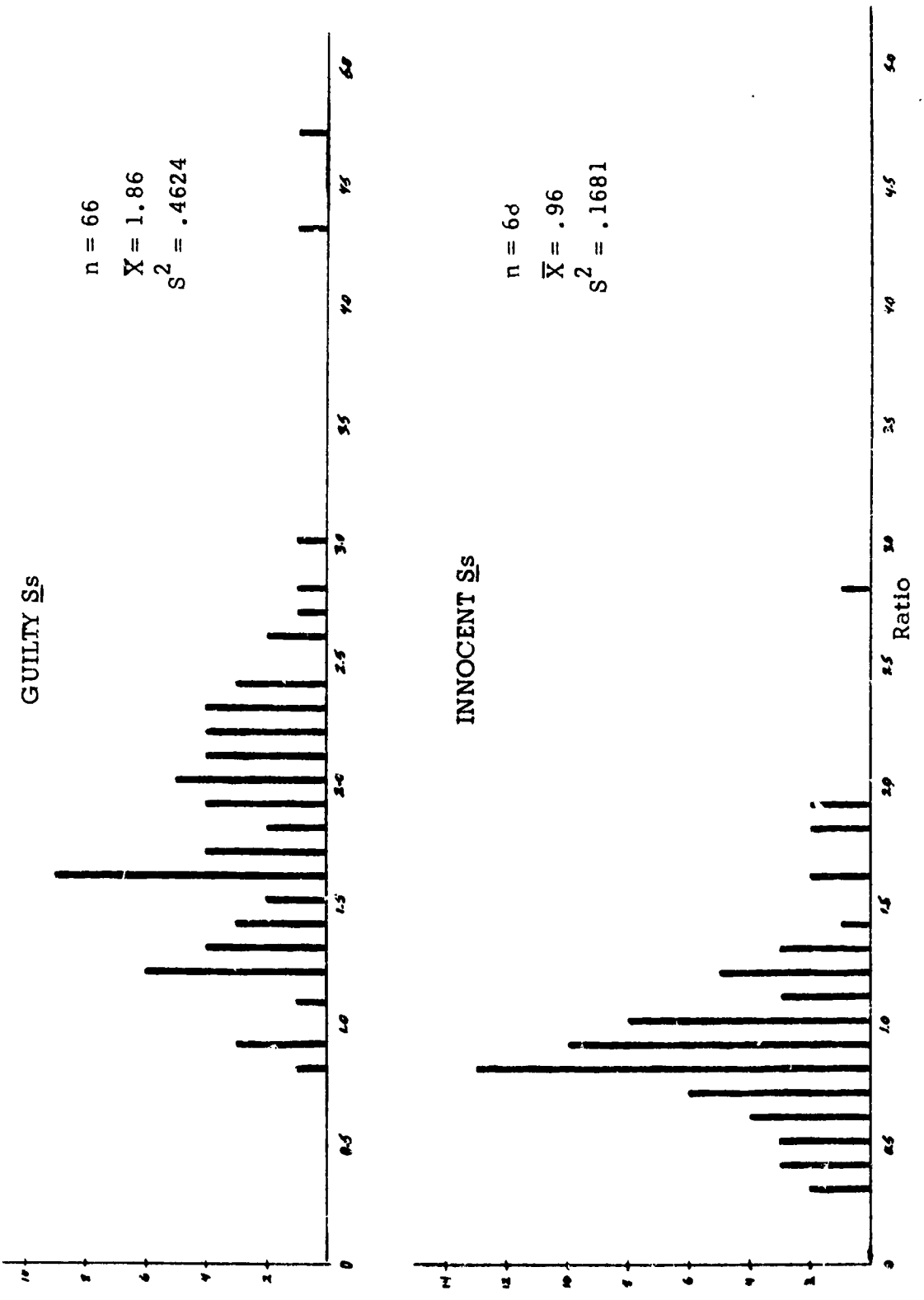

Figure 6. GSR ratio of critical to neutral words for all Ss in the RI series.

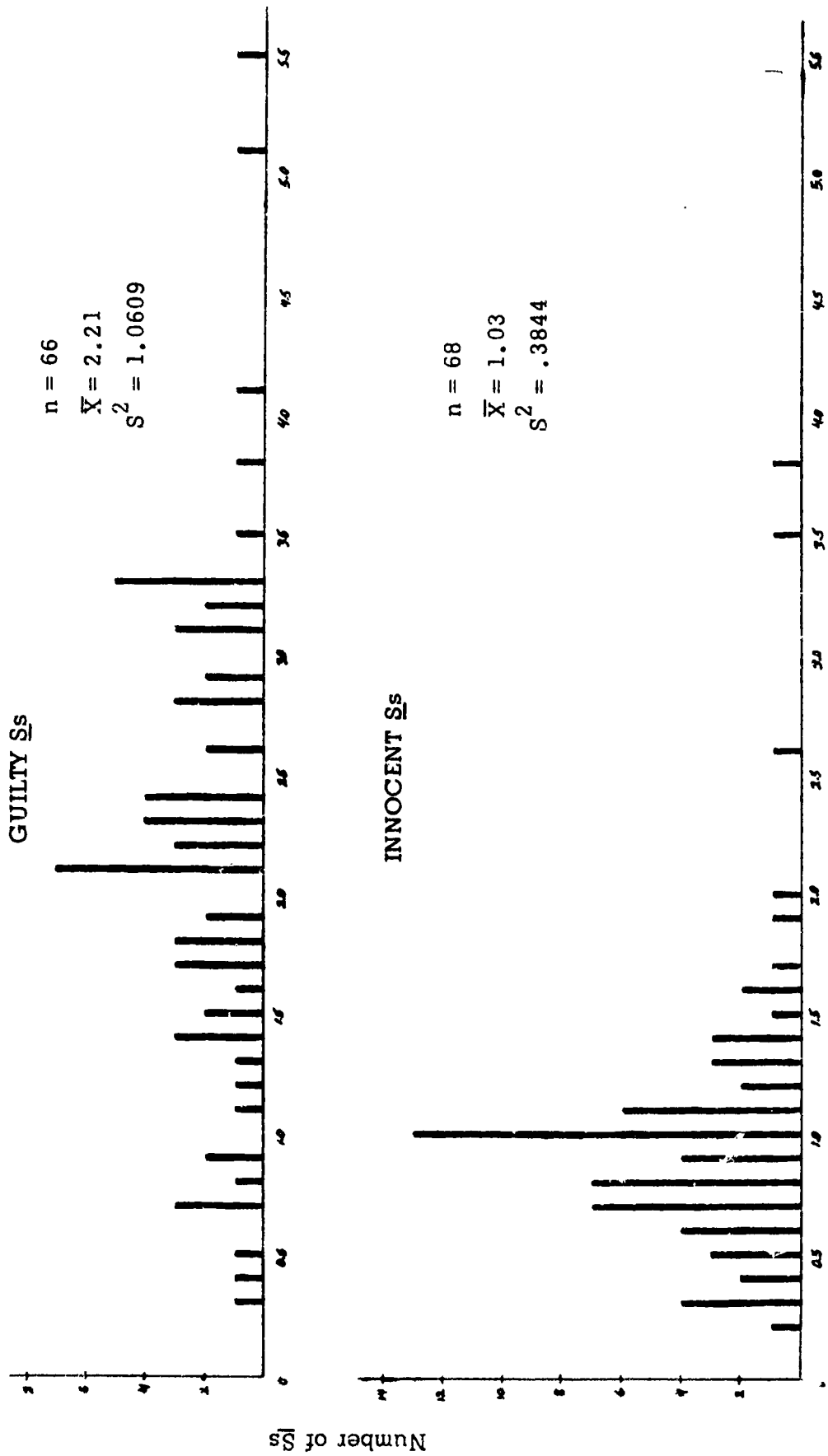

Figure 7. GSR ratio of the critical word to the mean of the neutral words for all Ss in the PT series.

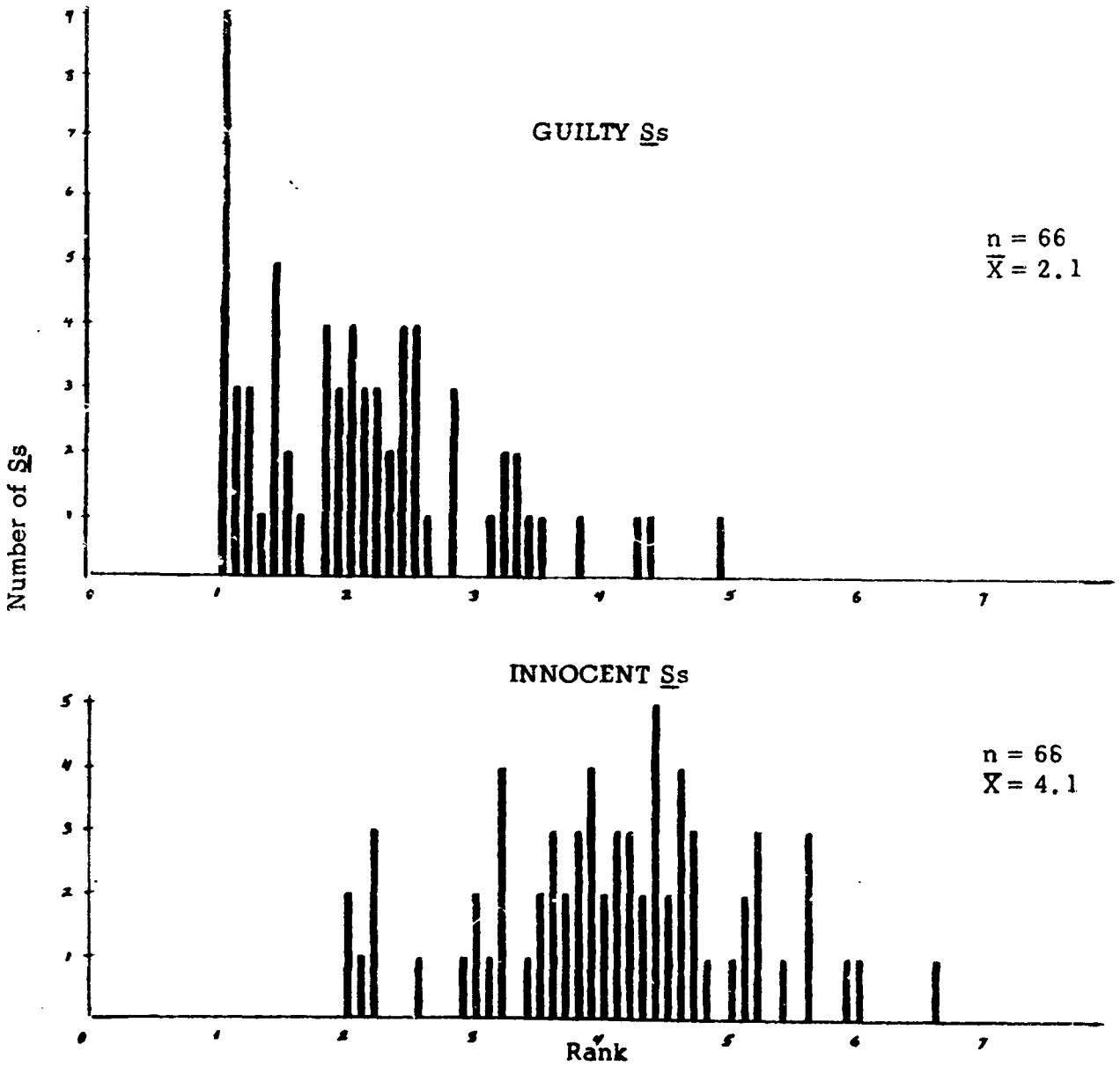

Figure 8. Mean GSR rank of the five critical words relative to the six neutral words in the RI series.

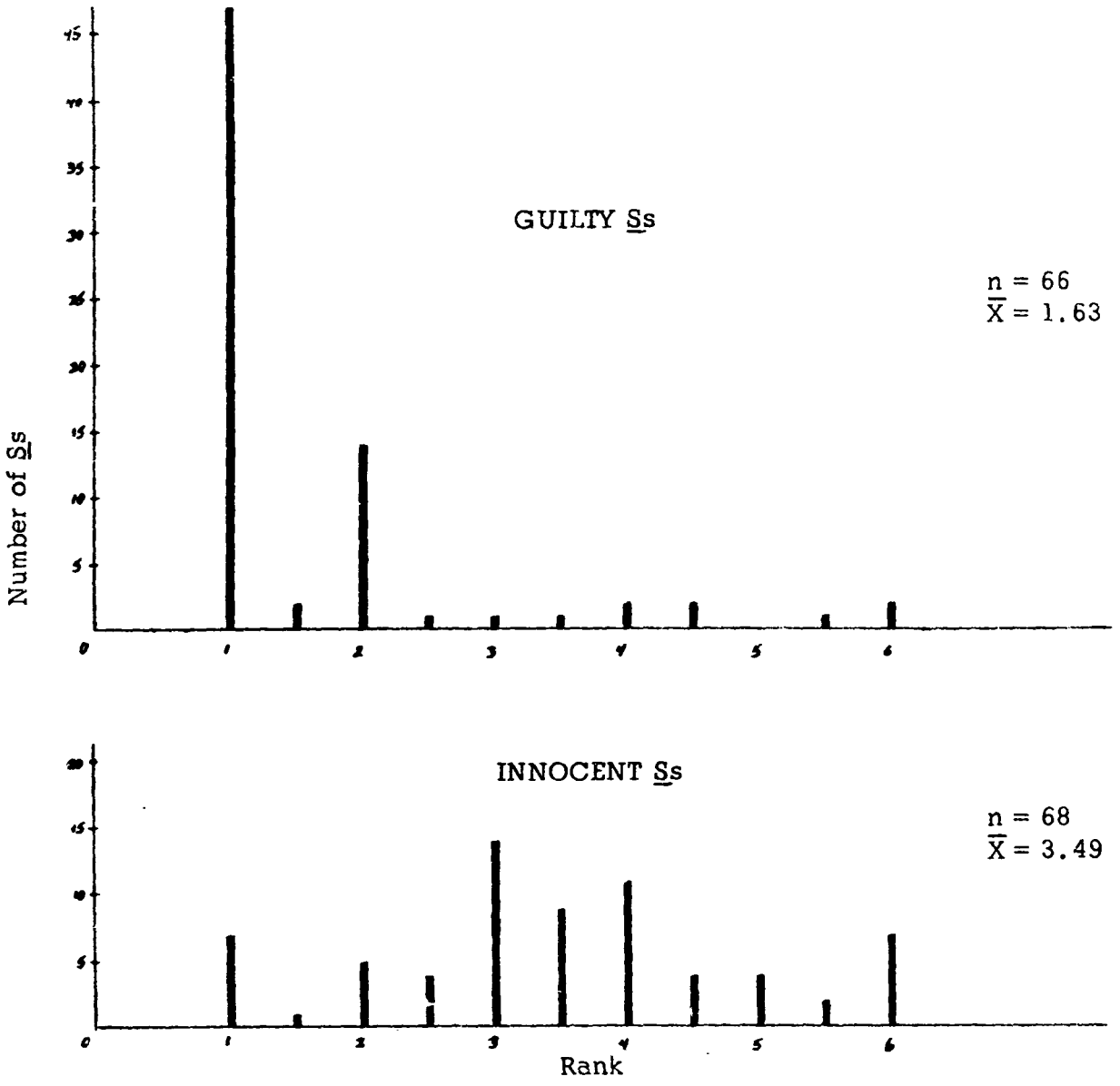

Figure 9. GSR rank of the color green (critical word) relative to the other colors in the PT series.

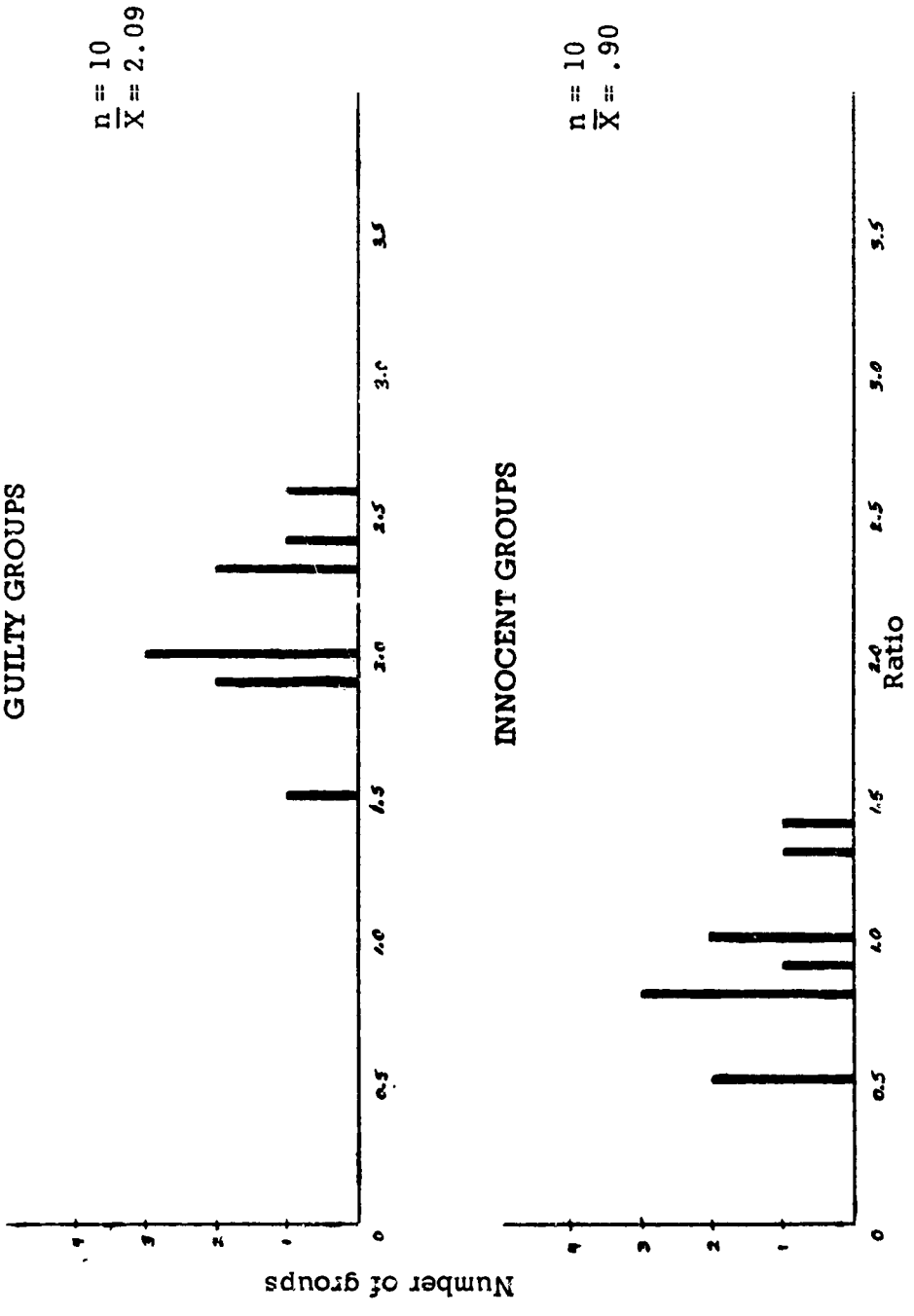

Figure 10. GSR ratio of critical to neutral words for all groups in the RI series.

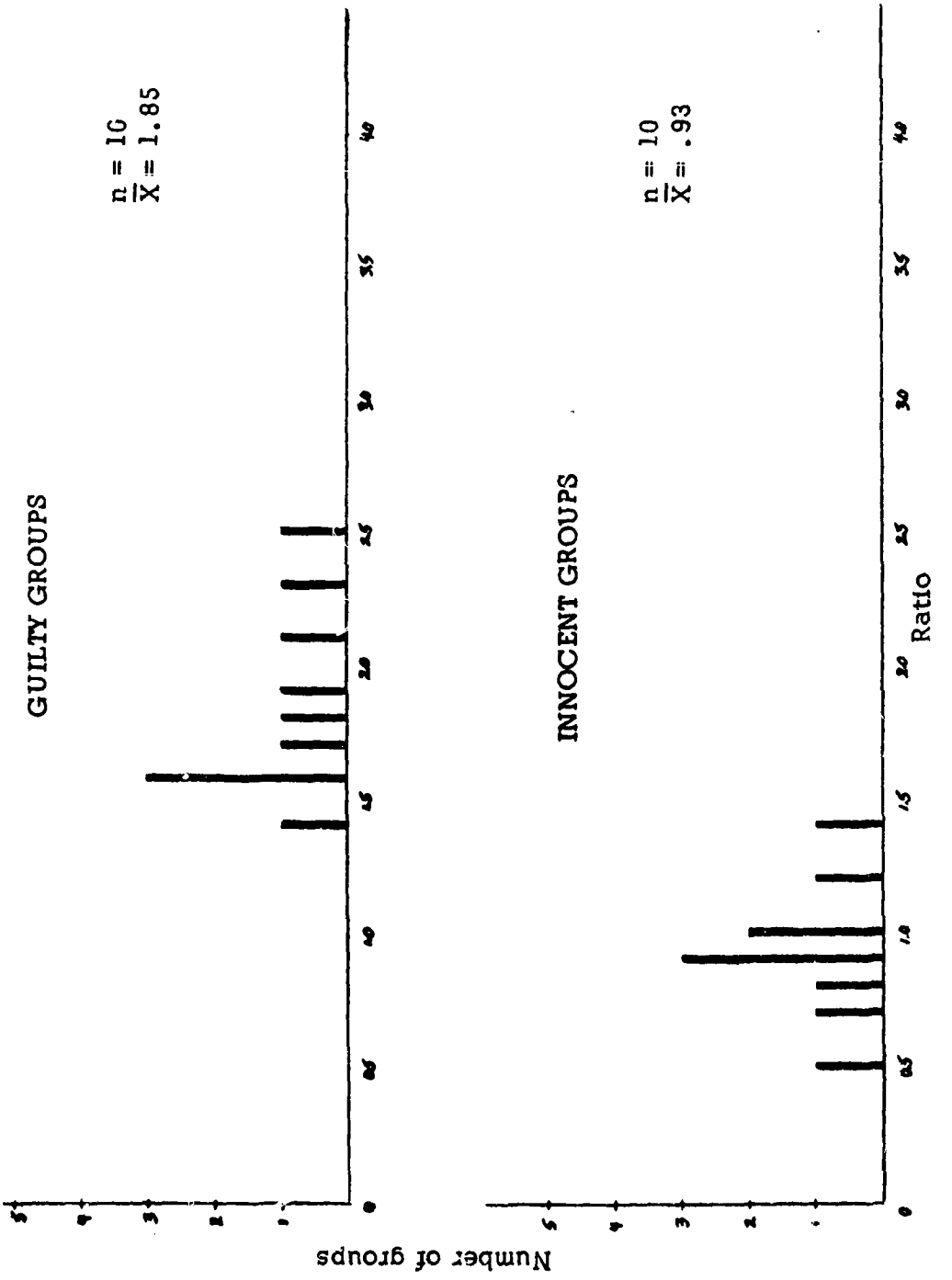

Figure 11. GSR ratio of the critical word to the mean of the neutral words for all groups in the PT series.

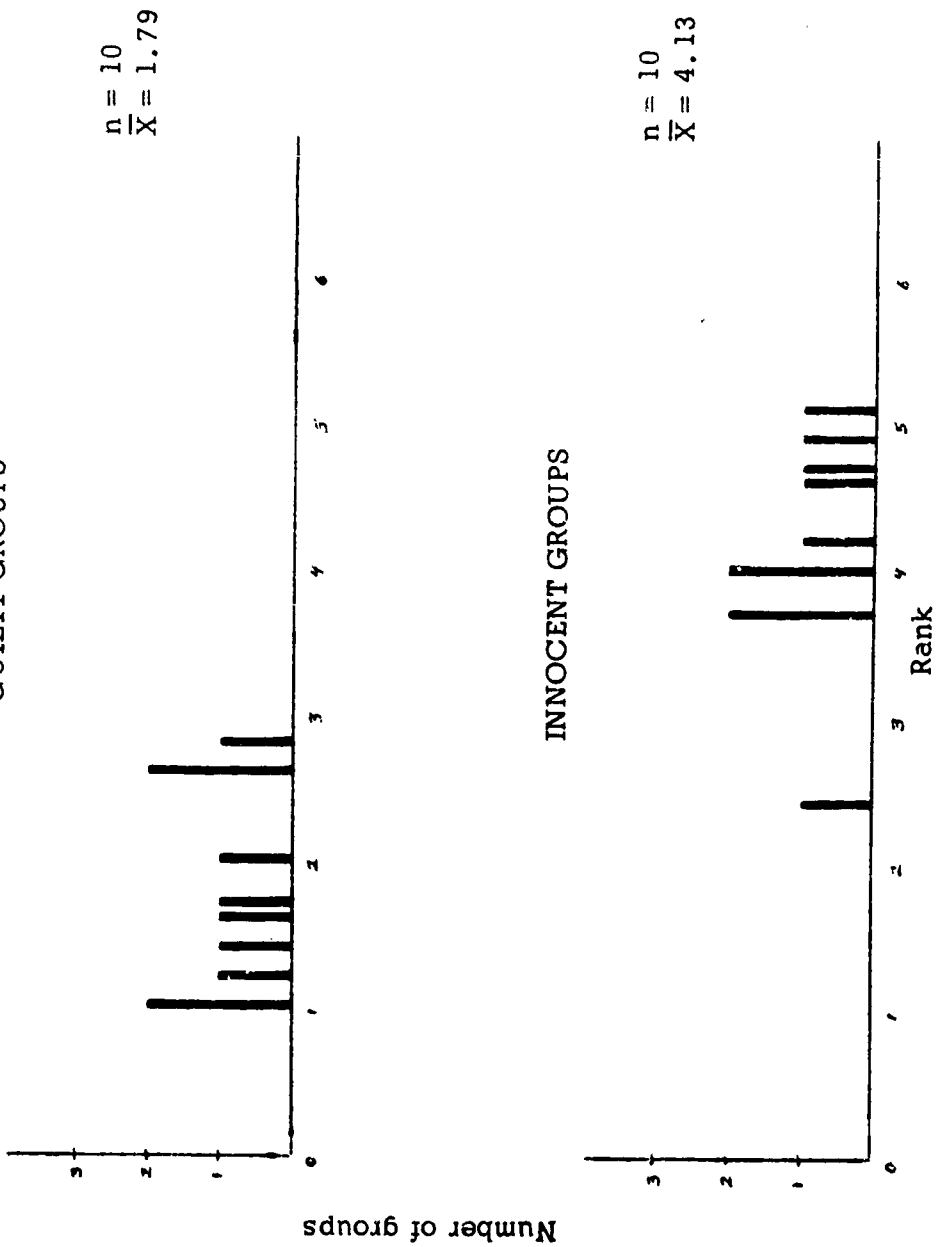

Figure 12. Mean GSR rank of the five critical words relative to the six neutral words for all groups in the RI series.

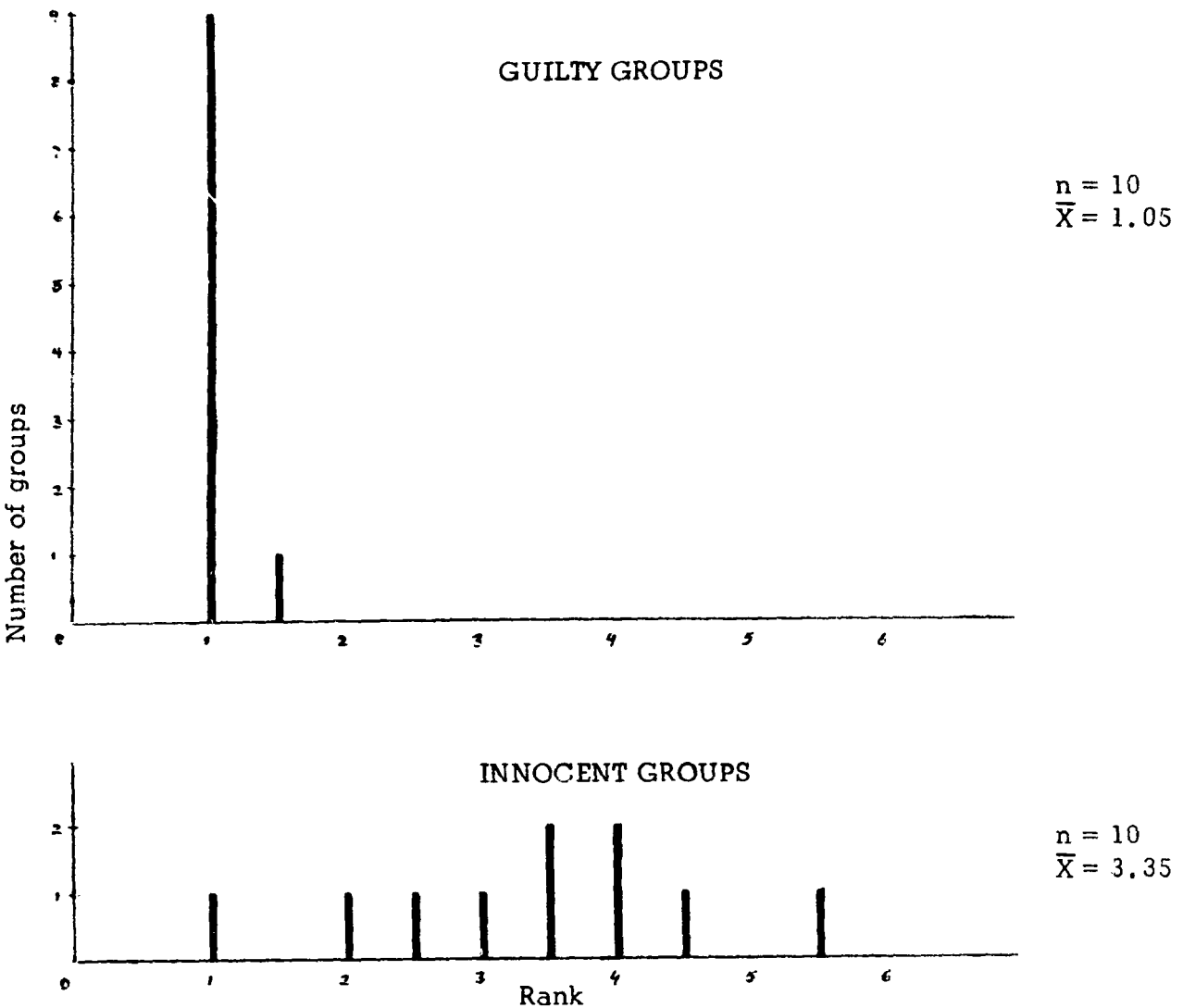

Figure 13. GSR rank of the color green (critical word) relative to the other colors for all groups in the PT series.

reveal rather clear separation. There is no overlap between the guilty and innocent distributions in the RI series, while in the PT series, the lowest guilty group ties with the highest innocent. When ranks are used, both the RI and PT series show good differentiation with little overlap. These results are presented in Figures 12 and 13.

DISCUSSION AND CONCLUSIONS

Previous research conducted by this laboratory on detection of deception(Gustafson & Orne, 1963; Gustafson & Orne, 1965) has clearly shown that the expectancies which a S brings to the laboratory play a major role in determining the ease with which he is detected. As in the above studies, Ss in the present experiment were initially informed that "while it is extremely difficult to beat the polygraph, people of superior intelligence and great emotional control are able to succeed." With Ss given this expectancy, it was predicted that those Ss in the guilty group who were shocked correctly for the cards they drew would perceive this as a failure on their part to deceive the polygraph, would be more strongly motivated to escape detection (as well as to avoid further shock), and hence be easier to detect during the actual interrogation. Conversely, it was predicted that Ss not receiving any shock for the cards actually drawn would perceive this as successful deception, would be less strongly motivated to deceive, and consequently be more difficult to detect. The group of Ss receiving shocks for cards which they did not actually draw were predicted to perceive the polygraph as detecting some response which they were emitting,

but a response unrelated to the card which they had drawn. This increased uncertainty would be expected to increase general arousal level (motivation) and hence increase detectability. It was further predicted that these Ss would be more detectable than those not receiving any shock, but be somewhat less detectable than the correctly shocked groups.

The results partially supported these predictions. When either ratio scores or ranks were used, the two shocked groups were significantly easier to detect than the nonshocked group in the PT series. In addition, further analyses conducted on the rank data revealed the SCC subgroup to differ significantly from the NS subgroup. No significant differences were obtained, however, between the two shocked conditions with either ratio scores or ranks, although when ranks were used, the Fisher Exact Probability value of .056 at least suggests that a significant difference between the two conditions might be achieved with larger samples.

When the RI data were analyzed, neither ratio scores nor ranks revealed any significant difference in mean response between the three subgroups. While the reason for the lack of significance in this series is unclear, it could be hypothesized that the general level of involvement of all three guilty subgroups was sufficiently high during this initial (RI) series to mask any differential subgroup effects which might have been present.

One further point of interest regarding the guilty subgroups is the significantly inflated variance of the ratio scores of the SICC Ss. This is present in both RI and PT series. It suggests that Ss within this

particular subgroup are not homogenous, but rather differ widely in their perception of the significance of the fact that they were incorrectly shocked during the initial part of the experiment.

In considering the overall distributions of the innocent and guilty Ss, both ratio and ranking methods appear equally effective as techniques for scoring the data. By choosing a cut-off point which yielded an equal percentage of overlap of guilty and innocent Ss, approximately 83 percent of the Ss were detected regardless of scoring technique or method of interrogation. This detection ratio is roughly comparable to that obtained by other investigators who have also used a "high involvement" condition (Kubis, 1962; Lykken, 1959). However, because of the many differences between these studies and the present one, it is difficult to make extensive comparisons of relative detection efficiency.

If cut-off points other than the one chosen were employed, it would be possible to virtually guarantee the absence of false negatives by including a larger percentage of false positives and vice versa. Although the separation of the guilty and innocent distributions was not as clear as was originally hoped, the technique of employing separate innocent and guilty distributions to obtain estimates of false positive and false negative rates with differing cut-off points is a feasible one and appears to warrant further exploration.

One of the most interesting findings in this study was the success achieved with the group GSR. In a conventional lie detection

situation where only a single individual is interrogated, emitting false positives to irrelevant items can be utilized to effectively mask response to critical information. Such countermeasures would be quite difficult to employ successfully when a number of individuals are simultaneously interrogated and a measure of group physiological reactivity obtained. Unless all individuals emitted false positives to the same items, which would be quite unlikely, these individual responses would randomize out across Ss. Consequently, this technique could have potentially useful applications in certain field situations in which a number of individuals possess some common items of information which they are unwilling to divulge.

EXPERIMENT II

The first experiment suggested the feasibility of the group GSR as a potentially useful technique in certain interrogation applications. In addition, the "courier" paradigm was shown to yield substantial separation of innocent and guilty individuals. However, it was felt that, although inclusion of the subgroups in the first experiment was of definite theoretical interest, the differential treatment of the subgroups made it difficult to interpret the effectiveness of the courier paradigm itself. Since an accurate evaluation of this paradigm was needed for future research, a second experiment without shock or threat of shock was designed. The following subsidiary issues were also explored.

1. What is the relative effectiveness of RI and PT methods when used in the context of this paradigm? (The first experiment employed only one critical word in a modified PT presentation, and hence did not allow an adequate comparison of the two methods.)

2. Conventional use of the PT technique involves the interrogator reviewing each of the questions with the suspect prior to the actual interrogation. It appeared desirable to approximate this in the laboratory situation and, in addition, to record physiological responses while the questions were being reviewed, with S less aware that this was actually part of the "interrogation." The question could then be asked: What is the physiological response to the same questions when the situation is structured as preliminary to interrogation as compared to the response to the "actual" interrogation? It was predicted that detection would be more difficult when S perceives the situation as preliminary to interrogation.

3. Could the findings of the first study relative to the effectiveness of the group GSR be replicated?

METHOD

Subjects

Twelve groups of 7 Ss each were solicited by advertisement in school newspapers. In order to maintain the same population used in the previous experiment, the ad called for Ss to participate in a "shock" experiment. All Ss were male college students between the ages of 18-23. Six groups were randomly assigned to the courier condition and six to the innocent condition. As with the preceding experiment, some Ss failed to report or arrived too late to be used. Consequently, the six courier groups contained a total of 38 Ss, while the six innocent groups contained a total of 35.

Procedure

The initial treatment of the courier and innocent groups was the same as that used in the previous experiment; that is, the courier groups spent their first hour learning the code words while the innocent groups were administered the attitude questionnaire. The particular code words learned, however, had not been used in the previous study and consisted of six words, with one selected from each of the following six categories: animals, trees, parts of the body, flowers, colors, and items of household furniture. All guilty Ss learned the same code words.

Following completion of their respective tasks, both innocent and courier groups were played a tape-recording telling them that they would now be interrogated by another individual (E_2). The tapes used were the same

as those employed in the previous study, except that for the courier groups, additional emphasis was placed upon the necessity of appearing completely innocent and not attempting to deceive the interrogator by trying to produce responses to irrelevant questions. These additional instructions seemed desirable, since in the first experiment it appeared that certain Ss were trying to escape detection by deliberately producing false positives.

E_1 then led the particular group to the interrogation room, Ss were assigned to chairs by E_2, and a tape was played which emphasized the polygraph's effectiveness in detecting lies. Upon completion of the tape, E_2 gave each S an opportunity to confess to being a courier. Following this an assistant attached the GSR electrodes, blindfolds, and dummy microphones to Ss.

The interrogation employed both RI and PT methods. Half the groups received RI first followed by PT, while the order was reversed for the remaining half. With the RI series, each of the critical words was embedded in separate blocks of six neutral words randomly selected from the six categories of words. Each successive block was presented without interruption. Placement of the critical item within each block was randomly determined, the only restriction being that it could not be either the first or last item in the block. The series of six blocks was repeated twice, with order reversed for the second presentation.

Prior to the RI phase of the interrogation, Ss were told that the words would be presented in a random order and that there was no way they could tell which word might be presented next. They were told to answer each

word with a yes or no, but to keep in mind that any response of yes would be an admission of being a courier. After the initial question, "Are any of the following words critical code words?", the words were presented to the Ss with interword intervals randomly varying from 8 to 15 seconds.

For the PT method, each block of seven words (six neutrals and one critical) consisted of words of the same category (colors, animals, trees, etc.). As with the RI series, location of the critical word within each block was randomly determined, with the same restriction that it could not be either the first or last item. With PT, however, it is necessary for S to know what questions are to be asked and their order of presentation. Consequently, prior to interrogating Ss on each block of words, the words were read to them at five-second interword intervals, and they were instructed to simply listen to the words. Instructions to the Ss emphasized that reading over the words prior to being questioned about them was simply to prevent any uncertainty or surprise regarding the words themselves or the order in which they would appear during the subsequent interrogation. Following the reading of words in each category, Ss were informed that they would now be interrogated about these same words. The words were then presented again with interword intervals ranging from 8 to 15 seconds.

All of the interrogation was presented to the Ss by tape-recording.

At the completion of the experiment, Ss were asked to write the six critical words if they were, in fact, a guilty group. The purpose of this was simply to insure that each guilty S knew the six code words.

Apparatus

With the exception of the shock apparatus which was not used in this experiment, all remaining apparatus was the same as that employed in the first study.

RESULTS

The data were scored in terms of ratio, difference, and rank scores. However, as with the first study, the different methods of scoring the data were found to yield essentially the same results. Consequently, only rank data are employed in the subsequent statistical analyses.

For the RI series, the GSR to the critical word in each block was ranked against the remaining neutral words. Mean ranks for both the first and second series of six blocks were determined separately for each S. In addition, the rank of the critical word in each alternate block of words in both series was determined and a mean for each S computed.

For PT, a somewhat different method of ranking was employed. While the RI method capitalizes on the "surprise" element, since S is completely uncertain as to when a critical question may occur, the PT method relies on S's anticipation of the critical item, resulting in a build-up in tension to the critical question followed by a subsequent release of tension. Consequently, with PT the GSR response to each word in a given block was subtracted from the GSR elicited by the preceding word. It was reasoned that if a build-up and release of tension did occur, the largest positive difference would be found between response to the critical word and response to the subsequent word. A rank of one was assigned to the largest positive difference and mean

ranks for both the initial reading and actual interrogation on the lists were separately determined for each S.

Relevant-Irrelevant vs Peak of Tension

In order to determine the relative effectiveness of these two methods of presentation, mean ranks on the alternate RI word blocks were compared with mean ranks on the interrogated PT lists. Table 1 shows the average ranks for both methods arranged according to order of presentation. Since it is obvious from inspection of these data that no order effects are present, both orders of presentation were combined. A Wilcoxon matched-pairs signed rank test was performed which yielded a z of 2.90 ($p < .01$, one-tailed).

Initial Reading vs Interrogation in PT Series

The mean rank for the two groups receiving the two orders of presentation on the initial reading of the lists was 2.31, while for the interrogation on the same lists the mean was 2.13. Although the difference is in the predicted direction, a Wilcoxon matched-pairs signed rank test yielded a z of only 1.30 ($p > .05$, one-tailed).

Distributions of Guilty and Innocent Ss

Both orders of presentation were combined for both guilty and innocent Ss and the distributions plotted separately. Since the RI method of presentation proved superior to the PT, only the RI data were plotted. These data are shown in Figure 1.

The method employed in the first study to determine a detection ratio was also employed with these data. The resulting ratio was 85 percent. This

Table 1

Mean rank of guilty Ss on critical words in the interrogated PT series and on alternate critical words in the RI series for both orders of presentation.

Order	Mean Ranks	
RI - PT	1.81	2.14
PT - RI	2.12	1.84

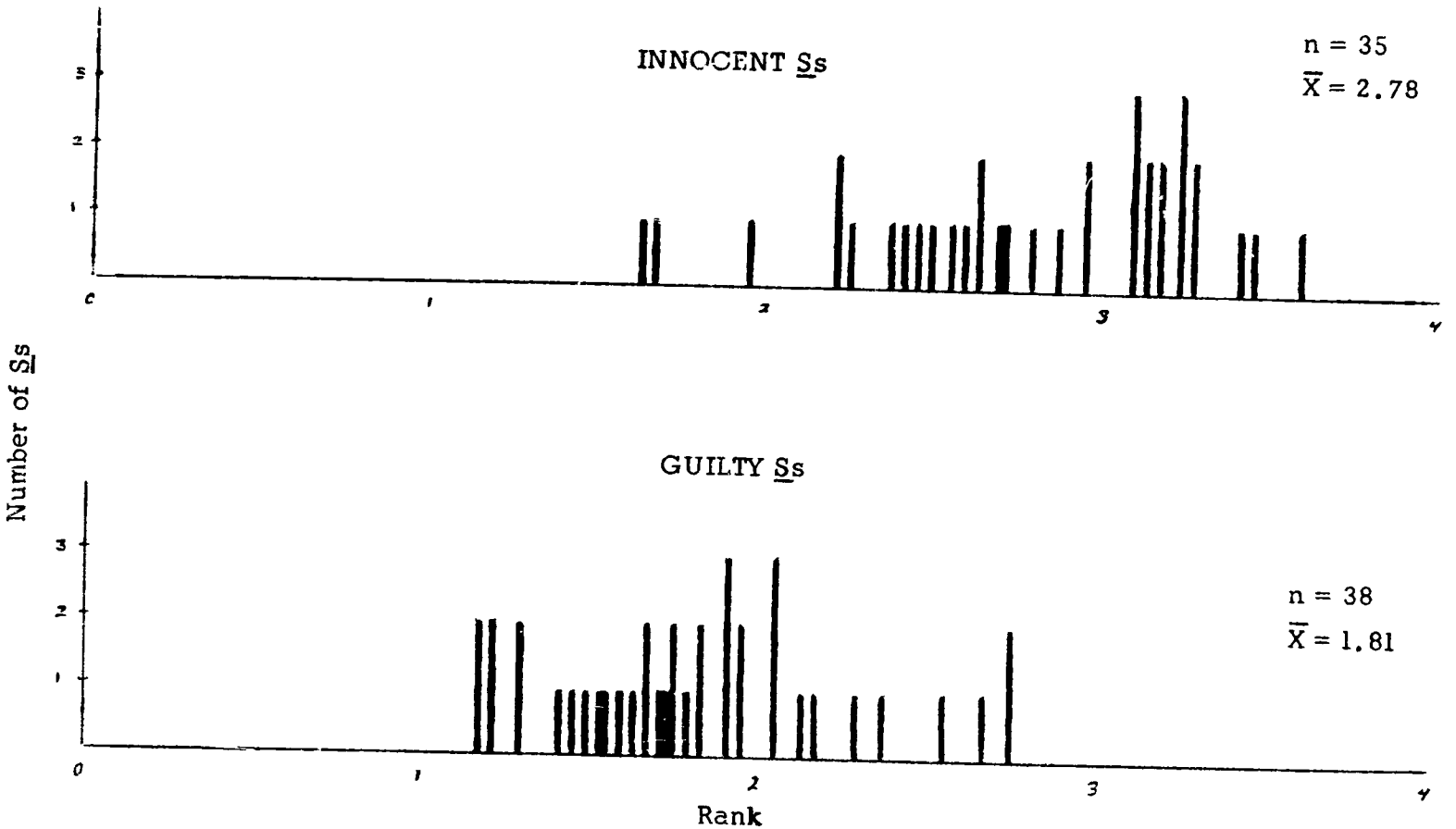

Figure 1. Mean GSR rank of the critical words for all Ss in the RI series.

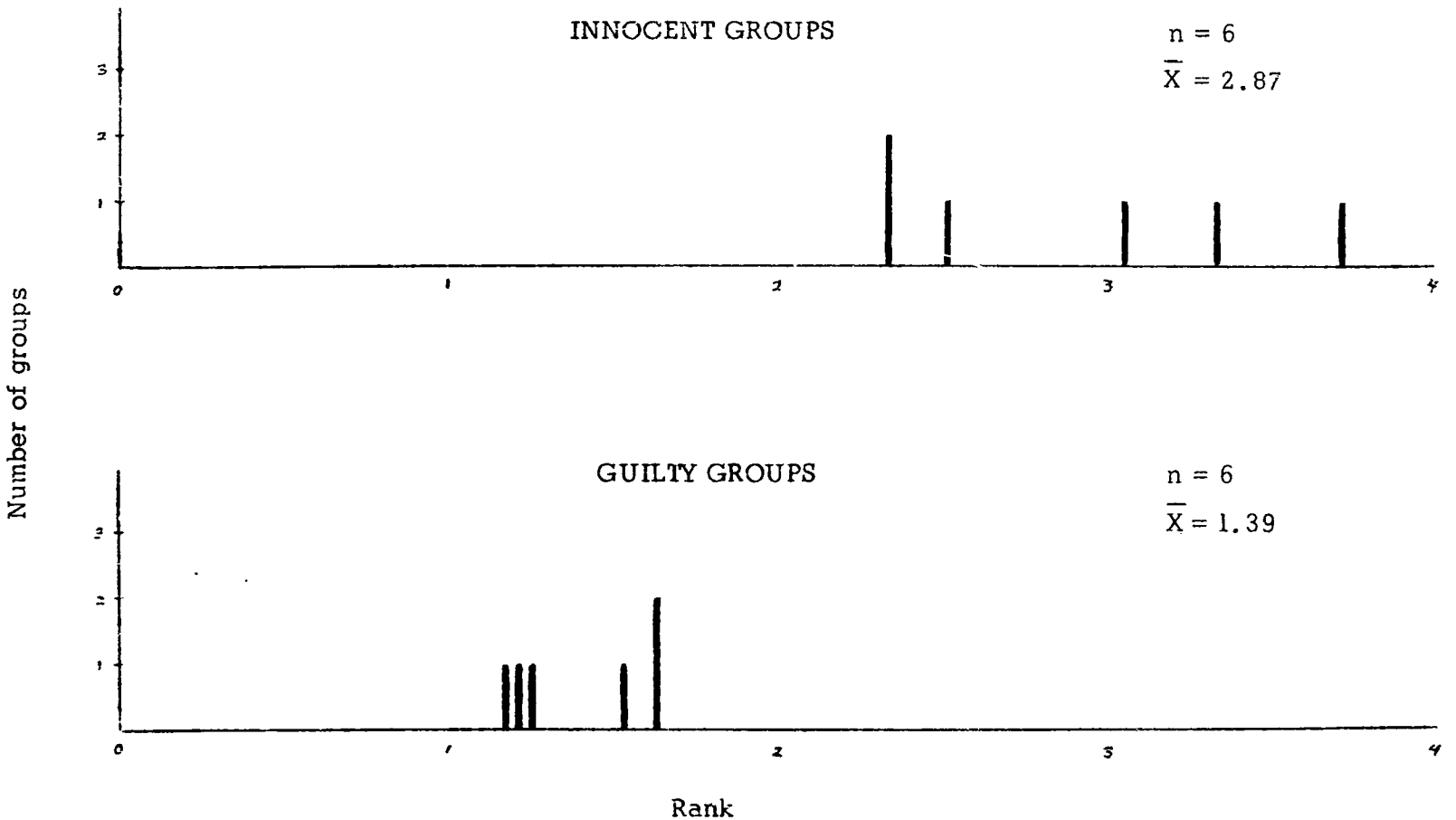

Figure 2. Mean GSR rank of the critical words for all groups in the RI series.

figure is approximately the same as that obtained in the previous study for comparable data, which indicates that despite the differences between the two studies, detection efficiency is about equal.

Group GSR

Figure 2 displays the group GSR data for the combined RI series. These data reveal clear separation of innocent and guilty groups with none of the overlap found in the individual distributions.

DISCUSSION AND CONCLUSIONS

A previous study conducted by this laboratory (Gustafson & Orne, 1964) demonstrated a definite superiority of the RI technique of interrogation as compared with the PT method, employing a "guilty information" paradigm. However, when a "guilty person" paradigm was used, the difference between the two methods of interrogation was no longer significant although the RI technique was still slightly superior in terms of mean ranks.

In the present study, which utilized a form of the guilty person paradigm, the RI method was found to be significantly superior to PT. The difference between the two methods obtained in this study, but not in the one cited above, may be a function of the many procedural differences between the two experiments.[4] Of these various differences, the most significant may have been the fact that Ss were tested in groups in the present study, while individual testing was done in the previous one. There was some observational evidence that a variety of subtle interactions

appeared to have taken place in the group context, and it may will be that, because of the group situation, Ss did not refrain from producing false-positive responses despite instructions to do so. This may have tended to reduce the effectiveness of the PT presentation, since it is somewhat easier to produce false-positives in PT than in RI because of S's knowledge as to when the critical items will occur.

It is possible, of course, that the superiority of the RI method shown in both experiments (significant in one experiment and not significant in the other), when the difference in procedures is ignored, is real. Thus, the surprise element inherent in the RI technique could be a more influential factor in the detection of deception than the build-up in tension characteristic of the peak-of-tension method, at least in the laboratory situation.

Concerning the question of whether a S can be detected when he is "unaware" that his responses are being evaluated, the results obtained for the two PT conditions are suggestive, but inconclusive. Although every effort was made to convince the Ss that the initial reading of the questions was preliminary to the actual interrogation, the fact that the GSR electrodes were attached during this period may have made the interrogator's arguments less than fully convincing. Nevertheless, the mean rank obtained from the "interrogated" condition exceeded the mean obtained when the questions were originally read and, although the difference was not statistically significant, it was

in the predicted direction. When one realizes that in reading over the questions to S, one is essentially employing the surprise element of the RI method and that the RI method is inherently superior to PT in these studies, the difference becomes more impressive. If the trend suggested by these data is substantiated in future studies, it would support the view that a S, unaware that he is being evaluated, may be difficult, if not impossible, to detect.

With regard to the group GSR, to fully explore the potential usefulness of this technique would, of course, require much more extensive investigation. The intent of both the present and previous experiments, however, was merely to explore the general feasibility of the technique itself. The data of the two experiments certainly suggests that this technique is an entirely feasible one and could have potentially significant applications.

It is striking that in both the present and previous experiment, the overall detection ratios were the same. Considering that in the one study shock, or threat of shock, was used, whereas this was not present in the second study, the stability of the detection ratio is particularly interesting. This would suggest that the involvement elicited by the courier paradigm itself accounts for a major portion of the variance.

In earlier studies (Gustafson & Orne, 1963; Gustafson & Orne, 1965), highly significant differences in detection rates were obtained through the use of motivating instructions and relatively minor, but

psychologically important, changes in interrogation procedures. The ease with which detection ratios were manipulated in these earlier studies make the stability of the ratio, in the face of the different experimental treatments employed in the two present studies, all the more striking. If one compares the earlier experimental situations with the present ones however, an interesting difference becomes clear. In the earlier work, the stimulus material was determined by selecting a card or some similar, obviously random, procedure. The interrogator himself offered the S a choice. Thus, the stimulus material was quite neutral and intrinsically unimportant to the S. Any significance the material had was a function of the instructions given by E. Under these circumstances, differences in the interrogation procedure and the instructions surrounding the procedure would be expected to yield very striking differences in detection ratios.

In the present two experiments, on the other hand, the stimulus material was taught by a different E and, while the material was initially of no consequence to S, the learning procedure was designed to strongly motivate S and make the stimulus material highly meaningful to him. Thus, before S entered the interrogation procedure, the stimulus materials had already acquired a considerable significance. Under these circumstances, the kind of experimental manipulations which were introduced into the interrogation procedure of the first experiment did not seem to have much effect on the detection ratios.

In view of these findings, it appears likely that changes in consequence during the interrogation procedure will have different effects, depending upon the degree of initial involvement with the stimulus material used. Thus, some kinds of manipulations may be additive, when one works with a high level of involvement, whereas others might not.

Involvement may be manipulated not only in terms of experimental context, as has been done in the past, but also by using stimuli which are known to have personal meaning, such as name or date of birth as opposed to inherently trivial stimuli, meaningful only in the context of the experiment. The possible interaction between the intrinsic importance of the stimulus to S and the involvement associated with the experimental context itself need to be investigated. It is also entirely plausible that the relative efficiency of different physiological parameters in a detection of deception experiment may depend upon the degree of involvement or arousal produced by different kinds of stimuli in different kinds of context. Future research will explore these issues.

References

Ellson, D. G., Davis, R. C., Saltzman, I. J., & Burke, C. J. A report of research on detection of deception. 1952. (Contract N6onr-18011 with Office of Naval Research.) Distributed by Department of Psychology, Indiana Univer., Bloomington, Indiana.

Gustafson, L. A., & Orne, M. T. The effects of heightened motivation on the detection of deception. J. appl. Psychol., 1963, 47, 408-411.

Gustafson, L. A., & Orne, M. T. The effects of task and method of stimulus presentation on the detection of deception. J. appl. Psychol., 1964, 48, 383-387.

Gustafson, L. A., & Orne, M. T. Effects of perceived role and role success on the detection of deception. J. appl. Psychol., 1965, 49, 412-417.

Kubis, J. F. Studies in lie detection: Computer feasibility considerations. RADC-TR-62-205.

Lykken, D. T. The GSR in the detection of guilt. J. appl. Psychol., 1959, 43, 385-388.

Lykken, D. T. The validity of the guilty knowledge technique: The effects of faking. J. appl. Psychol., 1960, 44, 258-262.

Footnotes

1. The authors wish to express their appreciation to Mary Jo Bryan for her assistance in the running of subjects and data reduction. Appreciation is also extended to Howard A. Keiser who served as E_1 in both experiments.

2. The blindfolds and headphones were employed in an attempt to reduce the amount of extraneous visual and auditory stimulation which could produce nonspecific GSR activity and hence interfere or mask GSR response to the critical items. The dummy microphone was used since each S was told that his verbal responses were being separately recorded.

3. It should be noted that in this and subsequent Figures, the subgroup n's differ slightly. This was due to errors made by E_2 during the card drawing, in that certain "correctly shocked" Ss were erroneously shocked for cards which they had not actually drawn. This resulted in a few groups containing less than two correctly shocked Ss.

4. The present study utilized a much more involving situation, employed different stimulus materials, tested

individuals in groups rather than singly, employed two different experimenters, and employed a different method scoring the peak-of-tension series.

APPENDIX

GUILTY Ss

Initial Instructions

In recent years there has been an increasing use of the polygraph as a tool to detect lying, both for the purposes of criminal interrogation as well as within the Federal government and private industry. This study is part of a set of experiments designed to test how well an individual is able to conceal information when examined by the polygraph. In this particular experiment we will be using a procedure based on an actual experience in the last war. The German Intelligence Service, in order to transmit messages, had a very interesting procedure. It employed two couriers. One of these couriers had the message itself, which was usually quite long and had to be written down. However, the message itself could not be decoded from the code book without having six critical words which would, as it were, give the particular key required to break that particular message. This set of code words was invariably sent by another courier who would not know the message. All he would know would be the six code words which had been committed to memory. The message could not be broken without these code words. The second courier, who had the six critical code words, did not have access to the message. Only when the two were put together could the message be readily deciphered. While it is true that the code could have been broken with a tremendous amount of work, time, and effort, time was usually of the essence, and it was essential that this be done as expeditiously as possible. Frequently we would manage to get hold of the message and have the code book, but we would not have the six critical

code words which were necessary to break the code for that message.

We would like to evaluate how the polygraph would have worked in this situation -- obtaining the six critical code words the courier had learned. In this experiment you are to play the role of someone who is to be interrogated. Let us suppose you had been to Europe recently and you are the courier. There would be a fair number of such individuals. You will be taught the critical code words, using an adaptation of the method used by the Germans to assure the couriers' knowledge and easy reproduction of these six critical words. At a moment's notice they were required to reproduce them while engaging in their normal tasks without any interference. At all times they would have them in the back of their minds. Obviously they were not written down because the slip might somehow get into enemy hands, but they had to be memorized thoroughly because it wouldn't be useful if the couriers arrived and had forgotten the code words. You know that a large number of people have been to Europe, and there would be no way for the interrogator to be certain that you are or are not a courier. Your task, obviously, is to appear to have no information, because if he in any way suspects that you are a courier, even if he cannot get the information itself, you would still be subjected to further intensive, unpleasant interrogation and would make delivery of the message difficult, if not impossible. So--your task is to learn these six critical code words. Then, during the interrogation, you must convince the interrogator that you know nothing--none of the words have any meaning and that about which he is asking you makes no sense at all. If he either gets the actual words which you are trying to hide or gets the

idea that you have any concealed information, you would be defeating your task in the experiment.

Of course the interrogator will attempt to cajole you into saying that you have some special information. He will attempt to have you, say--speak the truth! He will ask you to answer honestly. You will have to lie to him. You will have to convince him in every way possible that you are not a courier and do not possess the six critical code words.

Let me recap briefly. During the experiment you will learn six critical code words. You will have to know the words thoroughly--in all variations. You must not forget them. You will be tested on them repeatedly both before and after the experiment because, obviously, a courier who does not remember a message is useless. During the interrogation procedure itself, it is your task to convince the interrogator that you know nothing concerning these words about which he asks you. You will have to lie to him when he asks you to tell the truth. There will be other groups of subjects in this experiment who will be innocent. The interrogator will not know whether your group is guilty or whether it is innocent. Your task is to be guilty, but to convince the interrogator that you are innocent.

While this experiment will involve some electric shocks, you will not be shocked at the end of the experiment if you successfully convince the experimenter of your innocence.

One final word--it is possible for highly intelligent, mature, and stable individuals to control their emotions sufficiently to deceive the polygraph. That is exactly what we want you to try to do.

Good luck!

Instructions Given Following Learning of the Code Words

Gentlemen:

You have done very well thus far. You now know the six critical code words, and you know that you are to be couriers in this experiment. Shortly you will be interrogated, and the interrogator will do everything in his power to make you confess. Obviously no matter how much he tries to cajole you and talk you into confessing, you will not do so. It is your job to prove your innocence. It is your job to insist that you know nothing about these critical code words and to convince him that you are innocent, that everything which he asks you is strange and meaningless. This is a difficult task, but it is your job in the experiment. If you try hard enough you will be able to succeed at it. Remember he will try to talk you into confessing. Don't let him. He will try to pick up your responses on the polygraph. Try to hide them. Your job is to convince him of your innocence. Keep in mind that emotionally mature, intelligent subjects are able to succeed at it, and try your best. Good luck!

INNOCENT Ss

Initial Instructions

In recent years there has been an increasing use of the polygraph as a tool to detect lying both for the purposes of criminal investigation as well as within the Federal government and private industry. This study is part of a set of experiments designed to see how well an individual is able to conceal information when examined by the polygraph. In this particular study, we are testing the ability of individuals to deceive the polygraph examiner. There are some individuals in this study who will have certain guilty information. The study itself is based upon an actual experience in the last war. The German Intelligence Service habitually sent messages by two different couriers. One of the couriers would have the actual message written down, while the other would have the key to the code, which had no meaning of itself, but was essential to understand the message. We often found ourselves in the situation of having the actual code book, but not being able to solve the particular code without six critical code words which were being carried by the other courier. These code words would be committed to memory by the courier so that there was no message which could be found. While, of course, it was possible to break a code eventually, time was always of the essence, and it was essential that we try to get these words from the courier. In this experiment you are to play the role of an innocent individual who could be a courier, but who is not. Let us say that you have been to Europe recently and that it is possible for you to be a courier. You are under suspicion, but you are, in fact, innocent. Obviously your task is to convince the interrogator of your innocence because you do not enjoy being interrogated, you do not enjoy being detained, and you would like to get back

to normalcy. However, you are under suspicion, and the polygraph examiner will examine you. If you are able to convince the polygraph examiner that you are innocent, as you actually are, you will have done an excellent job in the experiment. That is what you are to do. Obviously, you will be somewhat anxious in this experiment as you would be in real life, but you simply do not have any information.

During the interrogation procedure you will receive some electric shocks. However, at the conclusion of the experiment and during the important part of the interrogation, you will not be shocked as long as your record appears innocent. This is again like real life. If you convince the interrogator of your innocence there will be no consequences.

Let me recap briefly. During this experiment you will be under suspicion of being a foreign courier, because of a recent visit to Europe. However, you are not a courier. You do not have any special information which you are trying to hide. You are innocent and your task is to convince the interrogator that you are innocent. This is what you are supposed to do.

Good luck!

Instructions Given Following Completion of the Attitude Questionnaire

During the next part of the experiment you will be interrogated. As you know some subjects in the experiment have information. They are couriers. You, however, have no such information. You are innocent. The interrogator will try to find out whether you are a courier. In doing so, he will try to threaten and cajole you into confessing. Obviously you do not want to confess because you are, in fact, innocent. In real life you would never confess, and in this situation you would not either. The interrogator will try to tell you that your guilt will show up on the polygraph. You have nothing to fear. It cannot show up on the polygraph because you are not guilty. So, do not let the interrogator frighten you into admitting something which is not so. Just don't be afraid of the procedure since you have nothing to hide.

Good luck!