CHAPTER 5

ASSESSMENT OF DETECTABILITY

This chapter reviews the results obtained in the two detection situations, the card test and the mock agent test, and describes the construction of indices of detectability. The data presented are for the electrodermal response, since this was the primary concern of the research programme. Preliminary analyses of some of the data revealed little capacity to detect deception using heart rate and respiratory measures and these were dropped from further consideration. This could well have been of course because of problems with these measures as discussed above (see p. 61-62). A more sensitive measure of the evoked cardiac response (e.g., heart rate deceleration) or the inspiration/expiration ratio or an amplitude measure in the case of respiration may have shown better results. There was also an attempt made to assess the records as might be done in a field situation. That is, activity in all three systems was scrutinized and a judgement made about deceptiveness. The results using this "clinical" approach as Szucko and Kleinmuntz (1981) have recently described it were equally disappointing and are not discussed further here.

The chapter is presented in two parts. The first part discusses the card test data, based on a total *n* of 121, and the second discusses the data for the 84 subjects participating in the mock agent test. Results are discussed both with respect to the sensitivity of the measures for the detection of deception (i.e., validity) and their psychometric properties (i.e., reliability).

Card Test

Results

Of the 121 subjects completing the card test 78 were presented with their chosen card during the test and 43 were not (one more than originally planned due to an oversight). The latter group constituted a control condition for determining the importance of actual choice of a card on differential autonomic responsiveness (see Chapter 3). Probably because the results with the card test have been so dramatic in individual cases, such a control has seldom been employed. Typically all subjects are presented with their chosen card, and the concern of the experimenter is in identifying which card was selected and not if one was selected. In the terminology of Orne et al. (1972) the card test typically is a test of guilty knowledge rather than a guilty person test. Inclusion of a control group permits an unequivocal determination of the role of choice in differential responsiveness. It also provides a sample of subjects for the assessment of false positive decisions, i.e., the proportion of cases likely to be detected as having selected a card when in fact they did not.

To provide a basis for comparison of the experimental and control groups, control subjects were randomly paired with 43 of the experimental subjects. For each pair, the card actually selected by the experimental subject was then taken as the critical or relevant card for the control subject. Only by chance would the control subject be expected to respond with a greater magnitude response to this card than to others. The fact of choice should, however, render the card "significant" for the experimental subject and hence enhance responsiveness to it. The pairing of subjects and the assignment of critical cards to control subjects in this way made possible a test of the discriminating power of the electrodermal measure.

Mean magnitudes of SCR to the non-chosen and the chosen cards for experimental and control subjects are presented in Figure 5.1. The plot is provided by condition (yes, no, and mute) and by trial, and reflects, except for the critical card, the order in which the cards were presented to subjects. Although the critical (selected) card was not always presented last, for convenience in plotting, response to it has been placed last in the series. The plot is for the 62 experimental and 40 control subjects for whom complete data for all points were available, since statistical analysis required the exclusion of all missing data. A separate plot using all subjects and averaging for missing values indicated that the conclusions to be drawn remained essentially unchanged.

Inspection of Figure 5.1 indicates that, in general, response magnitude decreases over card presentation, an habituation effect, but that this decline is greater on the first than on the second trial. Inspection of the figure further shows that response magnitude to the critical card was greater than to the others, but that this was true only for subjects in the experimental group. Third, inspection points to less responsiveness in the mute than in the yes or no conditions.

Statistical analysis of the data to check the reliability of these trends was performed using the BMD package (Dixson & Brown, 1981). In particular, BMDP2V was used to perform a repeated measures analysis of variance, first for the non-chosen cards, in which status (experimental/control) was a between groups factor, and serial position of the card (first through to fifth), trial (first or second), and condition (yes, no, mute) were within group factors. The analysis of variance summary table appears as Table 5.1. In assessing statistical significance the probability values provided by BMDP2V for the Greenhouse and Geisser (1959) correction were employed, since it could not be assumed that all the assumptions underlying the repeated measures analysis had been met.

Inspection of Table 5.1 indicates two significant three way interactions, one involving status, conditions, and trials, and the other involving conditions, trials, and serial position. Tests for simple main effects following (Winer, 1962) were therefore performed. In the case of the first interaction, analyses for the effects of status and conditions were performed separately for each trial. For trial 1, there was a significant main effect for conditions ($\underline{F} = 27.32$, $p < .001$). Comparisons of the means for the three conditions using the Newman-Keuls method indicated that under the mute condition responsiveness was significantly ($p < .01$) lower than under the yes or no conditions but that these conditions did not differ from each other. The effect for status was not significant ($\underline{F} = 1.18$), nor was the effect for the interaction of status and conditions ($\underline{F} = 1.16$). For trial 2, both the effects for status ($\underline{F} = 4.23$, $\underline{df} = 1, 107$, $p < .05$) and

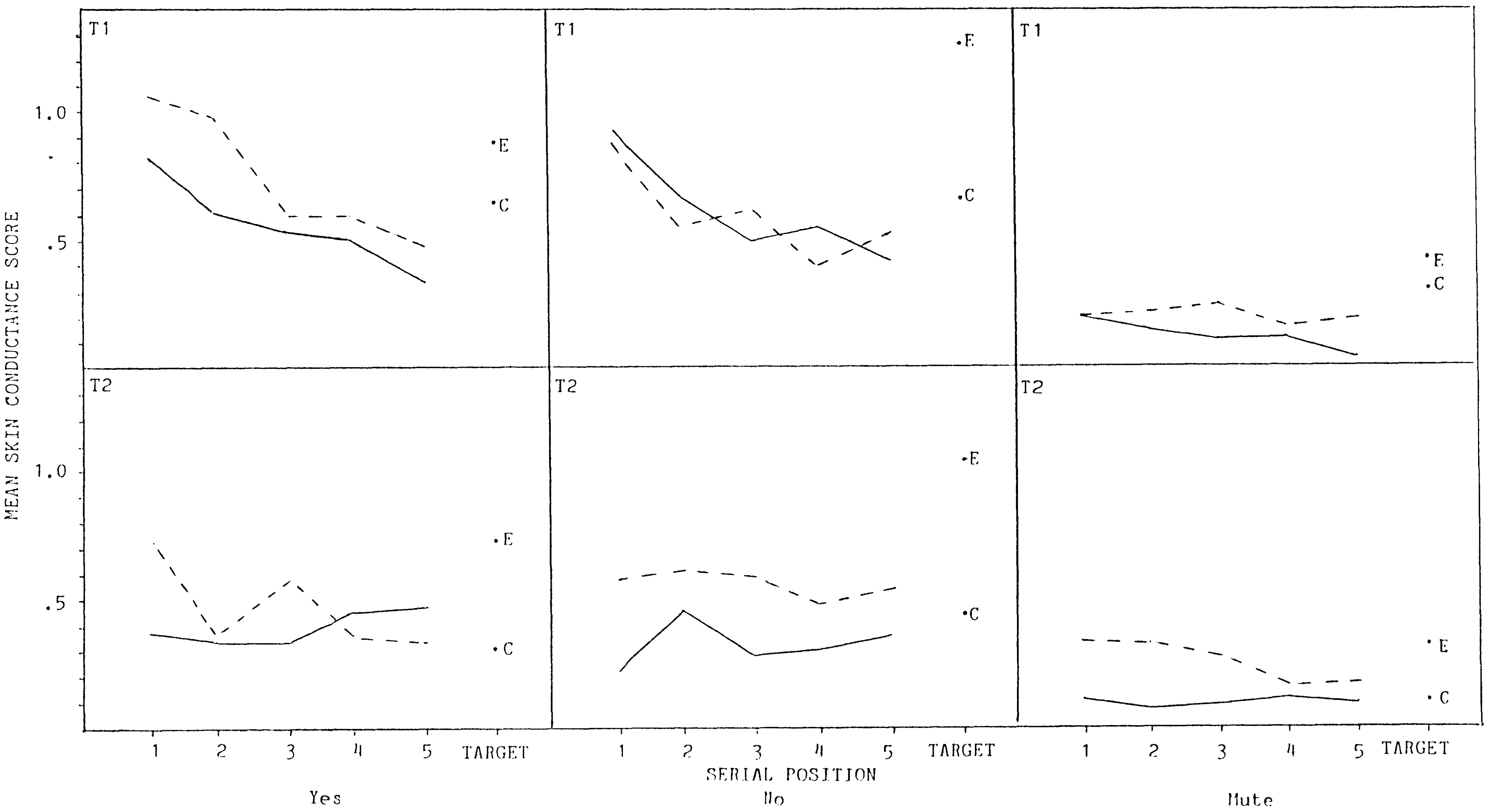


Figure 5.1. Mean skin conductance response during the card test for experimental (solid line) and control (dashed line) subjects. Panels left to right are for Yes, No, and Mute conditions respectively. Top panels are for

Table 5.1

Summary of Analysis of Variance on Magnitude of SCR to the Non-Critical Cards in the Card Test

Source of Variance	SS	df	MS	F
Guilt Status (S)	9.79	1	9.79	1.48
Error	662.51	100	6.63	
Condition (C)	87.82	2	43.91	30.52
C x S	.05	2	.03	.02
Error	287.75	200	1.44	
Trial (T)	13.59	1	13.59	18.53
T x S	1.00	1	1.00	1.36
Error	73.32	100	.73	
C x T	7.72	2	3.86	8.12
C x T x S	4.20	2	2.10	4.42
Error	95.17	100	.47	
Serial Position (SP)	15.26	4	3.82	11.00
SP x S	3.13	4	.78	2.26
Error	138.71	400	.35	
C x SP	3.87	8	.48	1.33
C x SP x S	2.40	8	.30	.83
Error	290.93	800	.36	
T x SP	7.22	4	1.80	5.93
T x SP x S	2.54	4	.64	2.09
Error	121.69	400	.30	
C x T x SP	6.97	8	.87	2.62
C x T x SP x S	1.88	8	.23	.71
Error	265.61	800	.33	

conditions (F = 20.06, df = 2, 214, $p < .001$) were significant, but the interaction was not (F = 1.51). Control subjects showed greater responsiveness than experimental subjects. The differences among conditions, again tested using the Newman-Keuls method, were the same as those occuring on trial 1.

Analysis of the conditions by trials by serial position interaction examined the effects for trials and serial position under each of the three conditions. For both the yes (F = 4.21, df = 4, 456, $p < .01$) and no (F = 6.07, df = 4, 436, $p < .01$) conditions but not for the mute condition (F = 0.29) the interaction of trials and serial position proved significant. Further analysis indicated that serial position was significant on trial 1 but not on trial 2 for both the yes (F = 8.71, df = 4, 456, $p < .01$) and no (F = 8.20, df = 4, 452, $p < .01$) conditions.

These analyses indicate (a) that conditions requiring a verbal response from the subject (yes or no) elicited substantially greater SC response than the condition in which the subject did not respond to the experimenter's questioning; and (b) that responsiveness decreased over presentations on trial 1 for those conditions in which a verbal response was required. The latter is interpretable as an habituation effect which is manifest only when a sufficient level of initial responsiveness is evoked.

To test for the effect of presentation of the critical stimulus on responsiveness, response magnitudes to each of the non-chosen cards were first averaged for each subject to provide a base-line against which to assess the impact of the chosen card. The comparisons are presented in Table 5.2.

Table 5.2

Means and Standard Deviations of SCR to Target and Non-Target Cards for Experimental and Control Groups in the Card Test

Response	Trial	Experimental				Control			
		Target		Non-Target		Target		Non-Target	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Yes	1	.90	1.18	.58	.79	.68	.79	.76	.84
	2	.72	1.17	.40	.63	.31	.60	.47	.65
No	1	1.27	1.70	.64	.86	.68	1.04	.61	.66
	2	1.04	1.44	.33	.55	.44	.94	.56	.74
Mute	1	.42	.94	.14	.30	.23	.51	.21	.25
	2	.32	.81	.10	.25	.20	.39	.25	.37

A repeated measures analysis of variance, again using BMDP2V, was conducted with the between subjects factor being status and the within subjects factors being condition, trial, and type of card (chosen/non-chosen). The results of this analysis are summarized in Table 5.3. Inspection of this table indicates significant main effects for all factors, except status, and significant interactions for status and card type, and condition and card type, and condition and trial. The first of these interactions can be seen from Table 5.2 to arise because the difference between response to the critical card and mean response to all other cards was greater in the case of experimental but not in the case of control subjects. The nature of the condition by card type interaction can be understood by summing response magnitude over groups and trials for each card type and for each of the three conditions. The difference between card types for each condition is greater for the no condition (1.29 micromho) than it is for the yes (.40 micromho) and the mute (.47 micromho) conditions which do not differ from each other. Thus the interaction is due to the significantly greater differential responsiveness to the critical card in the no condition. The interaction of condition and trial reflects the more marked decrease in response from trial 1 to trial 2 in the yes and no conditions as compared to the mute condition.

Table 5.3

Summary of Analysis of Variance Comparing Magnitude of SCR to the Critical and Non-Critical Cards in the Card Test

Source of Variance	SS	df	MS	F
Status (S)	4.35	1	4.35	.99
Error	436.04	99	4.40	
Condition (C)	46.53	2	23.27	24.80
C x S	2.61	2	1.30	1.39
Error	185.77	198	.94	
Trial (T)	7.83	1	7.83	23.24
T x S	.02	1	.02	.07
Error	33.36	99	.34	
C x T	2.73	2	1.37	5.33
C x T x S	1.03	2	.51	2.00
Error	50.74	198	.26	
Card Type (CT)	9.17	1	9.17	12.59
CT x S	15.69	1	15.69	21.54
Error	72.12	99	.73	
C x CT	2.90	2	1.45	4.08
C x CT x S	2.25	2	1.12	3.16
Error	70.46	198	.36	
T x CT	.21	1	.21	.92
T x CT x S	.25	1	.25	1.13
Error	22.01	99	.22	
C x T x CT	.01	2	.01	.03
C x T x CT x S	.20	2	.10	.48
Error	40.77	198	.21	

Following this overall analysis of the data, detectability indices were derived for each subject. The first of these, termed Differential Responsiveness (abbreviated subsequently as DIFR), was derived on the basis of the preceding analysis. That is, for each subject response magnitude averaged over the five non-chosen cards was subtracted from magnitude of response to the chosen card. Subjects (in the experimental group) for whom this value is positive can be described as correctly detected (i.e., response magnitude was greater to the chosen than to the non-chosen cards). Subjects for whom the difference is negative are not correctly detected, while no decision can be made for those for whom the difference is zero. These subjects might be considered undetected and classed with those with negative difference scores or as forming a separate category of "inconclusives."

Although based on the preceding analysis, this measure of Differential Responsiveness and the classification of subjects which it gives rise to are not strictly comparable with those of studies reported in the literature, since to derive the measure the experimenter must know which card the subject had selected. The card test, however, is typically, as in this study, administered blind and the chosen card is determined by comparing magnitudes of response among questions and taking as the chosen card that which elicits the greatest magnitude of response on questioning. To provide an index of detectability consistent with the use of such a strategy, the difference between response magnitude to the chosen card and the single largest response magnitude elicited by any other card was determined for each subject. Because it saw considered more comparable to other indices reported in the literature, this index was termed Detectablity (subsequently abbreviated DETEC). Subjects could again be classified using this index into those correctly detected (magnitude of response to the target card greater than that to any other card or positive difference), those incorrectly detected (negative difference) and the inconclusives (zero difference). The tables which follow summarize a number of features of the data observed using these two indices.

First, Tables 5.4 and 5.5 present the means and standard deviations for each index under each condition. Table 5.4 presents in a slightly different form data used in the preceding analysis of variance. Slight discrepancies between the means in this table and the corresponding values which can be calculated from Table 5.2 were brought about by the exclusion of cases with missing data in the analysis using BMDP2V. Inspection of Table 5.4 and 5.5 indicates the trends noted earlier: substantially greater differences for experimental than control subjects, and greater differences under the no than the other two conditions. Comparison of the two tables points to greater difference scores under the no condition using the index termed Differential Responsiveness than those obtained using the index termed Detectability.

Tables 5.6 and 5.7 present the results of intercorrelation of the indices across trials within conditions and across conditions averaged over trials. The purpose of these correlations was to provide estimates of the reliability of the indices of Differential Responsiveness and Detectablity, since success in prediction of an index depends in part on its reliability. Examination of Table 5.6 indicates that correlations for the control group with one exception are all close to zero whereas those for the experimental group are all statistically significant. This result is as would be expected since the difference score in the case of subjects in the control group should be the result of chance. In the case of subjects in the experimental group, however, the score should reflect the systematic effects of deception and hence show some reliability across conditions. As estimates of reliability, the correlations in the experimental group are disappointing. Even poorer, however, are the correlations for the experimental group in Table 5.7. In the case of this index, several of the correlations for the control group are statistically significant.

In view of the poor reliability of these indices, it was decided to average scores over trials to provide a more satisfactory set of indices for the study of relationships with measures of individual differences. Although averaging might have been done over conditions as well as trials, the significant difference between the no condition and the rest found in previous analyses suggested that conditions should be analysed separately. Means and standard deviations for the indices averaged over trials are presented in Tables 5.8 and 5.9. Regarding Table 5.9 it should be noted that the values in that table are not necessarily reproducible from Table 5.5. This results from the likelihood that a subject could have been detected on Trail 1 and not detected on Trial 2. When the total response to the cards is summed over trials the subject may or may not be detectable. That is, the outcome for Trial 1 and Trial 2 can be independent of the outcome over both trials.

Table 5.4

Means and Standard Deviations for the Difference between SCR to the Critical Card and Mean SCR to all Other Cards (Differential Responsiveness) on Each Trial of Each Condition of the Card Test

Condition	Trial	Experimental		Control	
		Mean	SD	Mean	SD
Yes	1	.28	.58	-.08	.96
	2	.25	.86	-.13	.59
No	1	.65	1.06	.17	1.07
	2	.59	1.17	-.11	.87
Mute	1	.28	.73	.00	.57
	2	.25	.74	-.07	.36

Table 5.5

Means and Standard Deviations for Detectability Scores on each Trial of each Condition of the Card Test

Condition	Trial	Experimental		Control	
		Mean	SD	Mean	SD
Yes	1	.21	.41	.12	.32
	2	.25	.43	.07	.26
No	1	.47	.50	.09	.29
	2	.41	.50	.07	.26
Mute	1	.30	.46	.12	.32
	2	.29	.46	.07	.26

Table 5.6

Intercorrelations of Indices of Differential Responsiveness across Trials within Conditions and across Conditions averaged over Trials in the Card Test

		Experimental[1]	Control[2]
Trials 1 versus 2	Yes	.25	.45
Trials 1 versus 2	No	.48	-.07
Trials 1 versus 2	Mute	.53	.11
Yes versus No	Averaged over Trials	.27	-.13
Yes versus Mute	Averaged over Trials	.27	.10
No versus Mute	Averaged over Trials	.42	.03

Notes: 1. For experimental group $r \geq .22$ to be significant at $p < .05$

2. For control group $r \geq .29$ to be significant at $p < .05$

Table 5.7

Intercorrelations of Indices of Detectability across Trials within Conditions and across Conditions averaged over Trials in the Card Test

		Experimental[1]	Control[2]
Trials 1 versus 2	Yes	-.07	.76
Trials 1 versus 2	No	.25	-.09
Trials 1 versus 2	Mute	.17	.47
Yes versus No	Averaged over Trials	.51	.43
Yes versus Mute	Averaged over Trials	.16	-.06
No versus Mute	Averaged over Trials	.04	.09

Notes: 1. For experimental group $r \geq .22$ to be significant at $p < .05$

2. For control group $r \geq .29$ to be significant at $p < .05$

Table 5.8

Means and Standard Deviations for the Difference Between the Mean Response to the Target Card and Mean Magnitude of Response to all Other Items (Differential Responsiveness) in Each Condition of the Card Test

Condition	Experimental		Control	
	Mean	SD	Mean	SD
Yes	.27	.58	-.11	.67
No	.62	.96	.03	.67
Mute	.25	.63	-.04	.35

Table 5.9

Means and Standard Deviations for Detectability Scores for Each Condition of the Card Test

Condition	Experimental		Control	
	Mean	SD	Mean	SD
Yes	.23	.42	.02	.15
No	.54	.50	.12	.32
Mute	.41	.49	.12	.32

Classification of subjects into those correctly detected, those incorrectly detected, and inconclusives is shown in Table 5.10 when the index of Differential Responsiveness is employed and in Table 5.11 when the index of Detectaility is used. The terms "correctly detected" and "incorrectly detected" as used in these tables with respect to subjects in the control group require some explanation. A control subject was classified as incorrectly detected if the difference score (Differential Responsiveness or Detectability) was positive. That is, the score indicated that the target card had been chosen, when of course it had not. By the same logic a control subject was classified as correctly detected if the difference score was negative. That is the decision rule is reversed for subjects in the control group as compared with subjects in the experimental group.

Inspection of Table 5.10 indicates that for subjects in the experimental group the frequency of correct detections in the case of the no condition is greater than that in either the yes ($\chi^2 = 5.47$, $\underline{df} = 1$, $p < .05$) or mute ($\chi^2 = 4.52$, $\underline{df} = 1$, $p < .05$) conditions. The difference is statistically significant as determined by a chi square test for dependent samples (Hays, 1963). This result is similar to that noted in earlier analyses of response magnitude: the no condition produced greater responsiveness to the target card. The same comparison in Table 5.11 is significant for the yes comparison ($\chi^2 = 5.35$, $\underline{df} = 1$, $p < .05$) but not for the no comparison ($\chi^2 = .13$, $\underline{df} = 1$, $p > .05$).

As far as accuracy rates for the control group are concerned, comparison of Tables 5.10 and 5.11 indicates that the Detectability index fares much better than the index of Differential Responsiveness. The apparent superiority of one index over the other must be considered, however, in terms of the operation involved in deriving them. To be correctly detected in terms of Detectability, response to any of the five non-target cards needs to be greater, for a control subject, than that to the target. That is, the probability is only one chance in six of meeting the decision rule and an extreme response on any of these will lead to the control subject being correctly detected in terms of this index. For Differential Responsiveness, on the other hand, an average is struck which reduces the influence of any extreme score and hence reduces the possibility of a correct detection. The probability in this case is one chance in two (either response to the target is

Table 5.10

Classification of Subjects in the Card Test in terms of Accuracy of Detection using the Index of Differential Responsiveness

Condition	Correctly Detected f	Correctly Detected %	Incorrectly Detected f	Incorrectly Detected %	Inconclusive f	Inconclusive %
Experimental						
Yes	43	56.6	23	30.3	10	13.1
No	55	70.5	12	15.4	11	14.1
Mute	41	53.2	21	27.3	15	19.5
Control						
Yes	21	48.8	15	34.9	7	16.3
No	22	51.2	13	30.2	8	18.6
Mute	23	53.5	10	23.2	10	23.3

Table 5.11

Classification of Subjects in the Card Test in terms of Accuracy of Detection using the Index of Detectability

Condition	Detectable f	Detectable %	Non-Detectable f	Non-Detectable %
Experimental				
Yes	18	23.1	60	76.9
No	42	53.8	36	46.2
Mute	32	41.0	46	59.0
Control				
Yes	42	97.7	1	2.3
No	38	88.4	5	11.6
Mute	38	88.4	5	11.6

larger or it is not) rather than the one in six with the index of Detectability. The "accuracy rates" for control subjects reflect these probabilities. This difference between the indices is also reflected in the absence of inconclusives using the index of Detectability, where the likelihood of the magnitude of response to the target card exactly matching that to all others is extremely small.

Discussion

The results of the analysis of variance of the SCR magnitude data point to three conclusions of some importance. First, SCR to the questions shows a pattern of habituation to the non-target cards and dishabituation or recovery to the target card. As such the invoking of OR theory in interpreting the results is appropriate, since habituation and dishabituation are primary defining features of the OR. Second, responsiveness is greater overall when the subject is required to make a verbal response.

The augmenting effect of linking a motor or verbal response to a stimulus in eliciting an OR is well accepted in the OR literature (Lynn, 1966; O'Gorman, 1977). Such an effect was particularly marked in the present data on the first trial. Third, the no condition leads to a greater differentiation of target and non-target cards than either the yes or mute conditions. That is, the content of the verbal response and not just the fact that a verbal response is required is significant in augmenting response. This point is taken up again below when the comparability of the present results with previous work is discussed.

One other conclusion can be drawn from the results of the analysis of variance, though this might be thought supererogatory by some. The conclusion is that selection of a card prior to questioning leads to enhanced response to it. This conclusion is based on the comparison of control and experimental groups and demonstrates in one sense the validity of the measures of detection derived. Without the use of a control group, strict logic does not permit the inference that it is the fact of prior selection per se which leads to increased responsiveness since presentation of the target card is confounded with serial position, nature of the card and the like. Although less than plausible that these confounds could account for the at times dramatic effects of presentation of the chosen card, the use of the experimental/control design as in the present study places the conclusion on a surer footing.

The effectiveness of the no condition in inducing Differential Responsiveness to the target card was seen in the analyses of the data which followed the overall analysis of variance. Whereas some 70% of subjects in the experimental group were correctly detected under the no condition when using the index of Differential Responsiveness only 53% to 56% were correctly detected under the other conditions. Correct detections were fewer when the index of Detectability was employed but again the superiority of the no condition was demonstrated. These results can be compared with those obtained in the two previous studies which examined the nature of the verbal response on detection in the card test. Gustafson and Orne (1956b) reported that 76% of subjects were correctly detected under the no condition as compared with only 58% under a mute condition, and argued as a consequence that the nature of the subject's response was critical. Kugelmass et al. (1967) disputed this conclusion as they found that the frequency of correct detections under a yes condition (70%) was in fact superior (though not significantly so) to that under a no condition (60%). The results of the present study are more consistent with the findings of Gustafson and Orne than with those of Kugelmass et al. Although the detection rates here are generally lower under all conditions to those reported in either of the other two studies, the major discrepancy is that between the yes condition here and that of Kugelmass et al. Table 5.12 presents the results of a series of chi square analyses which compared the present results with those obtained in the other studies under the relevant conditions. The only statistically significant effect is that between the yes conditions. The basis for this difference is hard to discern as the procedure followed here, except for the inclusion of a mute condition, was the same as that followed by Kugelmass et al. What the data do show is that while not a necessary condition for deception in the card test the act of saying No, i.e., lying on presentation of the critical card, enhances responsiveness to it. That is, a significant stimulus is rendered even more significant when there is competition between responses to be made to it.

Table 5.12

Chi Square Comparisons of Detection Rates for Comparable Conditions in This and Two Other Studies with the Card Test

Comparison	χ^2 Value
Yes condition (Kugelmass et al, 1967)	19.68^*
No condition (Gustafson & Orne, 1965b)	3.8401^1
No condition (Kugelmass et al, 1967)	$.24^{ns}$
Mute condition (Gustafson & Orne, 1965b)	2.23^{ns}

Note: 1. Strictly not significant as χ^2 (df = 1) must be greater than 3.841 at $\alpha = .05$ for significance.

* $p < .0005$

ns not significant

Although detection rates were broadly comparable to those obtained in previous studies, the reliabilities estimated for the indices were disappointingly low if the usual guidelines for psychometric tests are used as a reference point. These suggest that values of .9 or better are necessary for individual decision making and values of at least .6 for research purposes (e.g., Nunnally, 1967). Systematic investigation of the topic of reliability of indices of detection has not been reported previously and hence it is not possible to determine whether the estimates obtained here are representative or peculiar to the present sample. Two considerations suggest, however, that the results may not be atypical. First, the measures are difference scores which are notoriously unreliable (e.g., Linn & Slinde, 1977) since they compound the unreliability in each of the two measures which are used to derive the difference. Second, the reliabilities are essentially those for one-item tests which in psychometric work would be expected to have low validity. Instruments on which the guidelines cited above are based employ large numbers of items (e.g., 30+) and rely on the aggregation over items each of which is itself unreliable to bring about acceptable levels of reliability. Determining detection on one trial is the equivalent of a one item test. By averaging over two trials the reliability should in terms of psychometric theory improve, with the increment specifiable in terms of the Spearman-Brown Prephecy formula (Nunnally, 1967). When applied to the values in Tables 5.6 and 5.7 the formula can be used to indicate the reliability to be expected for a test containing twice as many items (i.e., based on two trials rather than one). Estimates derived in this way are presented in Table 5.13 which brings together a number of pieces of information relevant to interpretation of the indices derived from the card test. Only positive correlations have been corrected using the formula, since a negative value is best interpreted as indicating no reliability whatsoever for the index. Correction is therefore meaningless. Even with correction, only a few of the indices for the experimental group reach even minimal levels of reliability in the psychometric sense, fewer in fact than in the control group for whom the indices are wholly lacking in validity.

Until further research is directed to the problem of reliability, it is necessary to reserve judgement on the findings obtained here. What they do point to, however, is the constraint under which attempts to predict individual differences in terms of these indices must operate. Where measures do not correlate within themselves highly, high correlations with independent measures cannot be expected.

Table 5.13

Summary of Detection Indices derived from the Card Test

Index	Description	Detection Rate Exp. (n=78)	 Control (n=43)	Reliability Exp.	 Control
DIFRA1	Mean response to target card minus mean response to all other cards in the yes condition	57%	49%	.40	.62
DIFRA2	Mean response to target card minus mean response to all other cards in the no condition	71%	51%	.65	-.07
DIFRA3	Mean response to target card minus mean response to all other cards in the mute condition	53%	54%	.69	.18
DETCA1	Total response to target card minus largest total response to non-target card in the yes condition	23%	98%	-.07	.86
DETCA2	Total response to target card minus largest total response to non-target card in the no condition	54%	88%	.40	-.09
DETCA3	Total response to target card minus largest total response to non-target card in the mute condition	41%	88%	.29	.64

Mock Agent Test

Results

The mock agent procedure involved a CQ test and a GKT. Results for each of these are presented in turn.

The CQ test involved three relevant questions and two control questions (cf. p. 57-58), which were presented twice. The magnitudes of SCR to the relevant questions and to the control questions were averaged separately for each of the two trials for each subject. The means of these averages appear in Table 5.14. Inspecton of the table indicates that the average response to relevant questions was greater than that to control questions on both trials, although the difference on trial 1 was more marked. To test for the statistical significance of these effects, a repeated measures analysis of variance (BMDP2V) was performed on these data with question type (relevant/control) and trials as within subject factors. The results of this analysis are summarized in Table 5.15. There were significant main effects for question type and for the trials factor, as well as a significant interaction between question type and trials. The significant interaction can be attributed to the decrease in differential responding from trial 1 to trial 2 apparent in Table 5.14.

To provide an index of detectability for each subject for this section of the mock agent test, the difference between mean magnitudes of SCR to the relevant and control questions were calculated and the mean of the difference scores across the two trials found. The means and standard deviations of these scores appear in Table 5.16. To provide an estimate of reliability, difference scores for trials 1 and 2 were intercorrelated. The resulting coefficient (.21) although statistically significant ($p < .05$) is low when considered as a reliability coefficient. Combining the two scores was found, however, to produce some slight improvement in accuracy of detection as inspection of Table 5.17 indicates. This table presents the frequencies of subjects classified as "correctly detected" (positive difference: mean magnitude for relevant questions greater than that to control), "incorrectly detected" (negative difference), and "inconclusive" (zero difference) on the basis of difference scores for the first and second trials and for the two trials combined. Accuracy is greater on trial 1 than on trial 2 but is somewhat better for the composite index. For

Table 5.14

Means and Standard Deviations for SCR Magnitude to Relevant and Control Questions on Trials 1 and 2 of the Mock Agent Test

Trial	Question Type	Mean Response	SD
1	Relevant	.99	.72
1	Control	.55	.57
2	Relevant	.78	.70
2	Control	.53	.63

Table 5.15

Summary of Analysis of Variance on SCR Magnitude to Control and Relevant Questions in the Mock Agent Test

Source of Variance	SS	df	MS	F
Question Type (Q)	10.01	1	10.01	39.23
Error	21.17	83	.26	
Trial (T)	1.19	1	1.19	7.62
Error	12.94	83	.16	
Q x T	.75	1	.75	5.49
Error	11.34	83	.14	

Table 5.16

Means and Standard Deviations for Differences Between Mean Response to Relevant and Control Questions on Trials 1 and 2 and for the Trials Combined in the Mock Agent Test

Score	Mean	SD
Difference between mean response to relevant and control questions on Trial 1	.19	.54
Difference between mean response to relevant and control questions on Trial 2	.02	.47
Difference between mean response to relevant and control questions averaged over trials	.22	.78

Table 5.17

Detection Rates Employing Difference in SCR Magnitude to Relevant and Control Questions in the Mock Agent Test

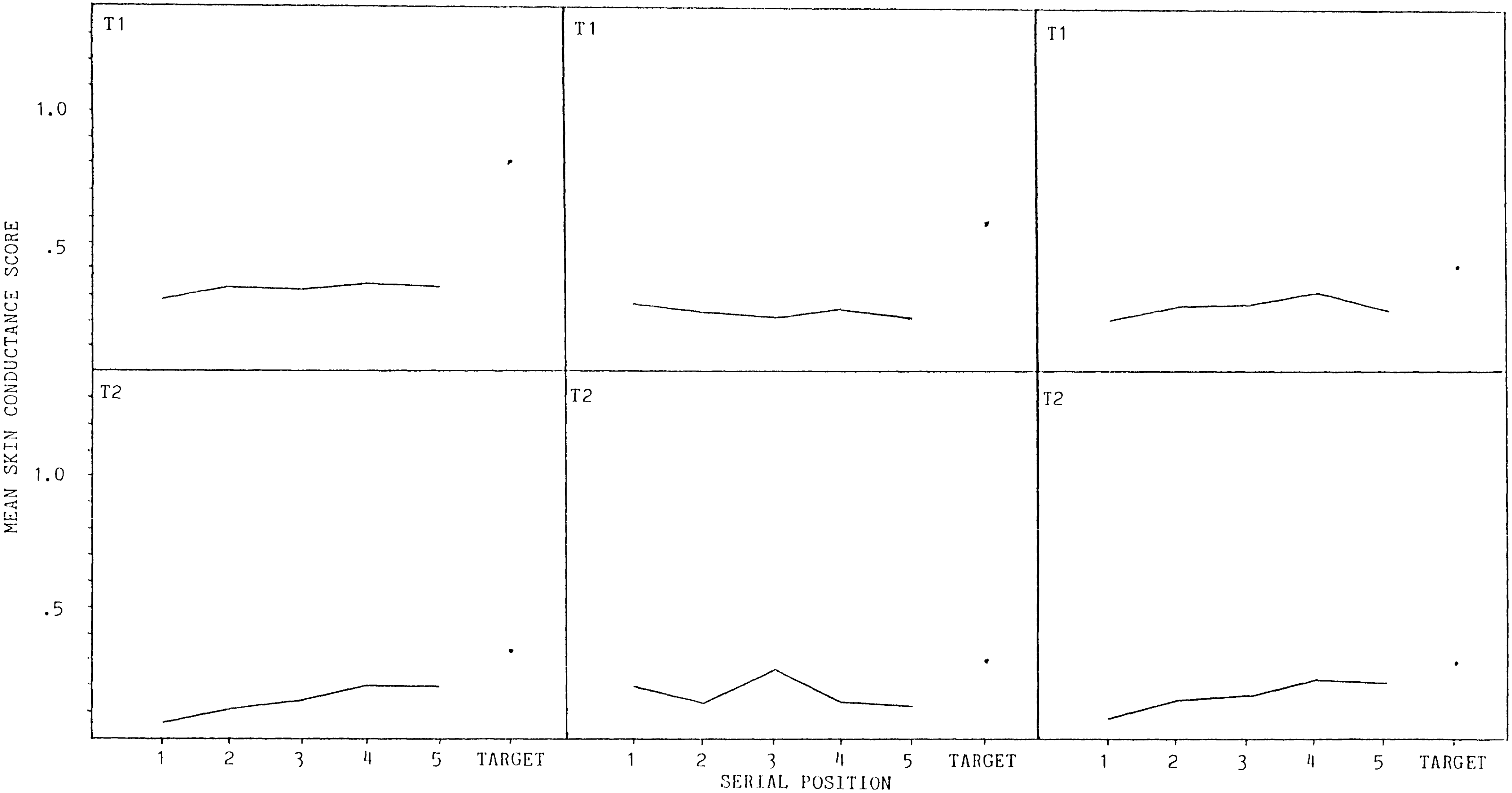
Trial	Correctly Detected		Incorrectly Detected		Inconclusive	
	f	%	f	%	f	%
First	53	63.9	29	34.9	1	1.2
Second	48	57.1	27	32.1	9	10.8
First and second	54	65.1	28	33.7	1	1.2

this reason, the composite was taken as the criterion of detectability for subsequent analyses.

The GKT involved presenting the subject with a total of 18 questions on each of two trials. On the first trial the subject was required to say No to each question, and on the second no verbal response was required. The questions differed in terms of the topic (birds, trees, and colours), and whether or not the subject had actually seen the item referred to in each question. Mean magnitude of SCR to each question is presented in Figure 5.2 for each trial. Mean magnitude to the critical item (i.e., the item the subject had seen) is graphed as the final point in each series, although this was not always its actual position in the series.

An initial analysis of the data in Figure 5.2 was performed on response magnitude to the questions involving non-critical items. The analysis, using BMDP2V, included three within subject factors, trials, topic (bird/tree/colour), and serial position. Results of the analysis are summarized in Table 5.18. With the conservative degrees of freedom proposed by Greenhouse and Geisser (1959) only the effect due to trial is statistically significant. That is, response magnitude to all questions was lower on the second trial when no response was required.

The second analysis involved comparing response magnitude to the critical item with mean magnitude of response to the non-critical items. The data for this analysis are summarized in Table 5.19. Examination of this table suggests that response to the critical items was consistently greater than to the non-critical item. The effect was, however, more marked on the first than on the second trial. There is also a trend for response to the critical items of the topic bird to be greater than that to critical items under the topic tree or colour. An analysis of variance was conducted on these data in which trials, topic, and item type (critical/non-critical) were all within subject factors. The results are summarized in Table 5.20. Although the main effects for trial, topic, and item type were all statistically significant, these effects must be interpreted in the light of the significant interactions in which the factors were involved. The significant trial by topic interaction could be inferred from the data presented in Table 5.19 as arising from the somewhat smaller differences between the two trials for the topic of colour than for the other two topics. Breakdown analyses



Bird
Tree
Colour

Figure 5.2. Mean skin conductance response to each question of each trial of the guilty knowledge test. Panels left to right are for bird, tree, and colour topics respectively. Top panels are for Trial 1 and bottom panels

Table 5.18

Summary of Analysis of Variance on SCR Magnitude to the Non-Critical Items in the GKT of the Mock Agent Test

Source of Variance	SS	df	MS	F
Trial (Tr)	6.35	1	6.35	16.49
Error	21.17	55	.38	
Topic (Tp)	.23	2	.12	.54
Error	23.73	110	.22	
Tr x Tp	.90	2	.45	3.63
Error	13.65	110	.12	
Serial Position (P)	.72	4	.18	2.21
Error	17.89	220	.08	
Tr x P	.36	4	.09	.88
Error	22.29	220	.10	
Tp x P	1.39	8	.17	1.45
Error	52.56	440	.12	
Tr x Tp x P	.51	8	.06	.53
Error	52.71	440	.12	

Table 5.19

Subjects' Skin Conductance Response and Standard Deviation to the Target Card and Mean Skin Conductance Response and Standard Deviation to Non-Target Cards in the Guilty Knowledge Test

Topic	Trial	Critical Items		Non-Critical Items	
		Response	SD	Response	SD
Bird	1	.81	.82	.37	.41
	2	.34	.90	.19	.29
Tree	1	.59	.74	.29	.36
	2	.31	.58	.21	.36
Colour	1	.42	.62	.28	.40
	2	.30	.53	.22	.35

Table 5.20

Summary of Analysis of Variance Comparing SCR Magnitude to the Critical and Non-Critical Items of the GKT in the Mock Agent Test

Source of Variance	SS	df	MS	F
Trial (Tr)	8.50	1	8.50	26.69
Error	22.92	72	.32	
Topic (Tp)	2.13	2	1.07	6.92
Error	22.16	144	.15	
Tr x Tp	2.05	2	1.03	7.06
Error	20.90	144	.15	
Item Type (I)	9.06	1	9.06	26.63
Error	24.50	72	.34	
Tr x I	1.87	1	1.87	15.74
Error	8.57	72	.12	
Tp x I	1.28	2	.64	5.40
Error	17.04	144	.12	
Tr x Tp x I	.49	2	.25	2.11
Error	16.86	144	.12	

indicated that the effect for trial was in fact weaker for colour (F = 4.30, df = 1, 81) than for either bird (F = 22.31, df = 1, 76) or tree (F = 26.85, df = 1, 79) but in all cases the effect was significant ($p < .05$). The trial by item-type interaction was due to the smaller difference between critical and non-critical items under the mute than under the no condition, but again breakdown analyses indicated that the effect was significant under both conditions (F = 42.96, df = 1, 76, $p < .001$ for no; F = 4.84, df = 1, 79, $p < .05$ for mute). The topic by item-type interaction was brought about by the larger difference between critical and non-critical items for the topic of bird than for that of colour (see Table 5.19). However breakdown analyses indicated that for all topics the differences were significant (F = 18.99, df = 1, 76, $p < .001$ for bird; F = 18.70, df = 1, 79, $p < .001$ for tree; F = 6.83, df = 1, 81, $p < .05$ for colour).

In interpreting these effects it must be noted that both the factors of trial and topic are confounded with an order effect. That is, trial 1 was always the trial on which a No response was required from the subject and trial 2 the trial on which no verbal response was required. Within each of these trials the topic bird preceded that of tree which in turn preceded that of colour. Order of presentation can be expected to exert considerable influence on physiological responses through the mechanism of habituation, and when confounded with other factors must be considered at least of equal explanatory value for any significant effects observed. In the case of the trial factor the confounding was intentional in that the GKT is typically used with a verbal no response and for this reason the no condition preceded the mute condition in the present study. For the factor of topic the confounding is more clearly a design fault. The result in both cases is the same: An unequivocal interpretation of the factors involved in the confound is not possible.

Following these analyses, indices of detection were constructed for each subject for use in subsequent analyses of individual differences. The indices were similar to those derived in the card test: Differential Responsiveness, which was the difference between magnitude of SCR to the critical item and mean magnitude of SCR across all non-critical items, and Detectability, which was the difference between magnitude of SCR to the critical item and magnitude of the largest SCR to a non-critical item. A similar set of analyses to those conducted on the card test indices were performed on these indices from the mock agent test and the results are summarized in the following tables.

Tables 5.21 and 5.22 present the means and standard deviations for two types of difference scores by topic and trial. There are slight discrepancies between the tabled values and those that can be calculated from Table 5.19 and Figure 5.2 since the latter arise from the BMDP2V analyses which involved the exclusion of any cases with missing values.

To provide an estimate of the reliability of these indices a series of intercorrelations were performed across topic areas for each trial separately. The results of this correlational analysis appear as Tables 5.23 and 5.24. For trial 1, at least, crrelations were higher for the index of Differential Responsiveness than for the index of Detectability. Although several of the coefficients are statistically significant, none are of large magnitude.

In order to increase reliability for subsequent analyses, indices were averaged. Since topic was not a factor of any theoretical interest, it was decided to average over topics within trials. The resulting mean for the Differential Responsiveness index was .288 (SD = .368) for trial 1 (the no condition) and .093 (SD = .380) for trial 2 (the mute condition). The corresponding values for the Detectability index were .048 (SD = 1.005) for trial 1 and -.452 (SD = .897) for trial 2. The correlation between trials was .48 ($p < .05$) for the Differential Responsiveness index and .21 ($p < .05$) for the Detectability index.

Finally, detection rates using each of the indices were derived using the decision rule employed with the data from the experimental group in the card test. Where a difference score was positive, the subject was classified as correctly detected (i.e., response to critical item greater than to non-critical items). A negative difference score was classified as an incorrect detection, and a zero was classified as inconclusive. Tables 5.25 and 5.26 summarize detection rates. It should be noted that inconclusives are far less likely when the Detectability index is employed because of the way in which this is derived (see earlier discussion of card test results).

Table 5.21

Means and Standard Deviations for the Difference Between SCR to the Critical Item and Mean SCR to all Other Items (Differential Responsiveness) for Each Topic and Trial of the GKT in the Mock Agent Test

Topic	Trial	Mean	SD
Bird	1	.41	.60
	2	.13	.81
Tree	1	.28	.52
	2	.09	.41
Colour	1	.13	.37
	2	.06	.45

Table 5.22

Means and Standard Deviations for Detectability Scores for Each Topic and Each Trial of the GKT in the Mock Agent Test

Topic	Trial	Mean	SD
Bird	1	.51	.50
	2	.18	.39
Tree	1	.36	.48
	2	.22	.42
Colour	1	.30	.46
	2	.16	.37

Table 5.23

Intercorrelations of Indices of Differential Responsiveness across Topics for Trial 1 (No Condition) and Trial 2 (Mute Condition) of the GKT in the Mock Agent Test

	Condition	
Comparison	No	Mute
Bird/Tree	.45	.21
Bird/Colour	.36	.03
Tree/Colour	.35	.24

<u>Note</u>: For experimental group $r \geqslant .22$ to be significant at $p < .05$

Table 5.24

Intercorrelations of Indices of Detectability across Topics for Trial 1 (No condition) and Trial 2 (Mute Condition) of the GKT in the Mock Agent Test

	Trial	
Comparison	1	2
Bird/Tree	.19	.22
Bird/Colour	.18	.14
Tree/Colour	-.10	.17

<u>Note</u>: For experimental group $r \geqslant .215$ to be significant at $p < .05$

Table 5.25

Detection Rates when the Index of Differential Responsiveness is used to Classify Subjects in the GKT of the Mock Agent Test

Condition	Correctly Detected f	Correctly Detected %	Incorrectly Detected f	Incorrectly Detected %	Inconclusive f	Inconclusive %
No	62	80.5	11	14.3	4	5.2
Mute	39	48.8	26	32.5	15	18.8

Table 5.26

Detection Rates when the Index of Detectablity is used to Classify Subjects in the GKT of the Mock Agent Test

Condition	Detectable f	Detectable %	Non-Detectable f	Non-Detectable %
No	44	52.4	40	47.6
Mute	23	27.4	61	72.6

Discussion

The results of the analyses of variance in both the CQ and the GKT sections of the mock agent test indicated that statistically significant discrimination between critical and non-critical items was obtained. Indices based on the difference in response magnitude to the two types of items can therefore be considered to have some validity. The actual detection rates obtained using these indices were, however, lower than those reported in the literature, a point that calls for some comment.

For the CQ test, studies by Waid (Waid, Orne & Orne, 1981; Waid, Orne & Wilson, 1979; Waid, Wilson & Orne, 1981) using a mock agent paradigm and listed in Table 1.3 report a mean accuracy rate of detecting guilty subjects of 76.5%. Studies by Raskin (Barland & Raskin, 1975; Podlesny & Raskin, 1978; Raskin & Hare, 1978) using a mock theft paradigm and listed in Table 1.3 report a mean accuracy rate of detecting guilty subjects of 90.3%. Both these accuracy rates are substantially higher than the 65% found here. In view of this, the detection rate found here cannot be considered typical of those reported in the literature for the CQ test.

The procedure followed here departed in several respects from that usually followed in administering the CQ test. These departures were described earlier but need to be considered again here. First, the examiner knew that the subjects in the mock agent test were all guilty. This, it must be conceded, may have produced some degree of bias in the examiner, but the extent of this bias and its influence on the results are unknown. Second, subjects were not motivated to deceive. High levels of motivation were purposely not induced because of the need to induce a range of detectability scores. Certainly, the literature is clear on the point that highly motivated subjects are more easily detected than non-motivated subjects (Gustafson & Orne, 1963; Raskin & Hare, 1978). Lower average rates of detection are therefore to be expected. Third, the list of questions that constituted the CQ test were read to the subject. While this is not a strict departure from the normal practice of the pre-test interview, it is possible that the standardized procedure followed for all subjects resulted in less emphasis being placed on the control questions than is usually the case. In addition the type of control question employed in the study was the 'non-exclusive' control question, which Podlesny and Raskin (1978) have shown to be less effective in identifying guilty subjects than the 'exclusive' type. All these departures could have contributed to the lower detection rates found here as compared to previous studies with the CQ test.

Better detection was observed in the GKT when the index of Differential Responsiveness was employed, at least for trial 1. Detection decreased markedly on trial 2 but whether this was due to the change in response requirement on the subject or an habituation effect cannot be determined because of the confounding noted above. The detection rate for trial 1 was 81% which compares favourably with that reported in Waid et al. (1979) of 53% and in Waid et al. (1978) of 79%. The actual index employed in classifying subjects cannot, however, be compared across the studies. When a more similar index to that employed in other studies (the index of Detectability) is used here for classification the detection rate drops to 52%. This was not improved when the data were re-analysed (results not reported here) using a numerical scoring procedure more similar to that employed by Lykken (1959). When the rates for trial 2 are considered, no comparable result with previous studies can be found and hence this index must be considered suspect.

As far as reliability of the indices is concerned, the same relatively low values found in the case of the card test were observed. The correlations when corrected using the Spearman-Brown Prophecy formula yield the estimates reported in Table 5.27 which also summarizes a number of other aspects of the indices derived from the Mock Agent test. The correction factor applied in obtaining these reliability estimates was, in the case of the CQ indices, 2 as the "test" can be considered as doubled when the average is over two trials and, in the case of the GKT, 3 as here the average was over three topics. The only index which meets even a minimal level of reliability in the psychometric sense is the index of Differential Responsiveness for trial 1 (No condition) of the GKT. The representativeness of these results cannot be determined because of the lack of attention to the problem of reliability in the literature.

Table 5.27

Summary of Accuracy Rates and Reliability for Indices Derived from the Mock Agent Test

Abbreviation	Description	Accuracy Rate n = 84	Reliability[1]
DRC12	Mean magnitude of response to relevant questions minus mean magnitude of response to control questions	65%	.35
DIFRA4	Mean magnitude to target cards minus mean response to all other cards in the no condition	81%	.66
DIFRA5	Mean response to target cards minus mean response to all other cards in the mute condition	49%	.36
DETCA4	Total response to target cards minus largest total response to a non-target card in the no condition	52%	.23
DETCA5	Total response to target cards minus largest total resonse to a non-target card in the mute condition	27%	.40

Note: 1. Reliabilities in the no and mute conditions are means across the bird, tree, and colour topics.

CHAPTER 6

TESTS OF HYPOTHESES

The previous two chapters have described the derivation of a number of indices of electrodermal activity on the one hand, and a number of indices of Differential Responsiveness and Detectability on the other. The present chapter examines the relationships between these two sets of indices and relationships with the self report tests of personality in the light of the hypotheses formulated in Chapter 2. The examination is in two parts. First, the bivariate correlations are examined between selected indices of electrodermal activity and personality and responsiveness in the various phases of the laboratory tests of deception. As there is potential overlap among a number of these indices, the second part of the chapter considers the multivariate or multiple correlations between the predictors and Differential Reponsiveness and Detectability. In these multivariate analyses the various self report measures of personality are studied separately. The conclusions to be drawn from these several analyses are summarized in a final section of the chapter.

Bivariate Analyses

Table 6.1 presents the product moment correlation coefficients between each of the indices of electrodermal activity and the indices of Differential Responsiveness derived from the card test. Correlations with the self report measures of personality are also included. Table 6.2 presents the correlations with the Detectability indices from the card test, and Table 6.3 presents the correlations with the Differential Responsiveness and Detectability indices derived from the mock agent procedure. Summaries of the ways in which each of the indices of Differential Responsiveness and Detectability were derived are presented in Tables 5.13 and 5.27.

Table 6.1

Correlations Between the Predictor Variables and the Differential Responsiveness Indices of the Card Test for Experimental and Control Subjects

	DIFRA1		DIFRA2		DIFRA3	
	Experimental	Control	Experimental	Control	Experimental	Control
SCBL	-.004	.011	.016	.221	.042	.093
SCAMP	-.023	-.133	-.025	.123	.000	.377**
NSR	.159	.101	.278**	.122	.395****	-.070
TNR	.161	-.093	.264*	-.041	.191	.068
IRSSCB	.334***	.065	.454****	.028	.441****	-.246
E	.007	-.081	-.144	.100	-.133	.008
N	.071	-.411***	.143	.158	.095	.341*
L	-.158	.089	.104	-.052	-.004	.099
P	.186	-.076	.266	.125	.183	-.120
So	-.052	.211	-.199	-.244	-.129	.048

Note: The correlations involving the electrodermal indices were calculated on an *n* of 75 for guilty subjects and an *n* of 43 for innocent subjects. The correlations involving the personality variables were calculated on an *n* of 39 for guilty subjects and an *n* of 43 for innocent subjects.

**** $p < .001$
*** $p < .01$
** $p < .02$
* $p < .05$

Table 6.2

Correlations Between the Predictor Variables and the Detectability Indices of the Card Test for Experimental and Control Subjects

	DETCA1		DETCA2		DETCA3	
	Experimental	Control	Experimental	Control	Experimental	Control
SCBL	.094	.252	-.000	.274	-.040	.307*
SCAMP	.105	-.038	-.019	.166	-.081	.426***
NSR	.121	.267	-.013	.145	.378****	.108
TNR	.242*	.058	.102	-.089	.370****	.257
IRSSCB	.177	.336*	.183	.138	.332***	.026
E	.098	-.180	-.193	-.054	.226	-.121
N	-.094	-.149	-.118	-.041	.100	.365**
L	.110	.092	.170	.083	.037	-.049
P	.136	.014	.143	.226	.214	.082
So	-.004	.131	-.292	-.097	.077	-.039

Note: For the _ns_ on which these correlations are based see footnote to Table 6.1.

**** $p < .001$

*** $p < .01$

** $p < .02$

* $p < .05$

Table 6.3

Correlations Between Predictor Variables and the Detectability Indices of the Mock Agent Paradigm ($\underline{n}$ = 73)

	DIFRA4	DIFRA5	DETCA4	DETCA5	DRC12
SCBL	.330***	.230*	.153	.248*	.057
SCAMP	.513****	.226	.251*	.215	.253*
NSR	.307***	.192	.066	.290**	-.004
TNR	.541****	.280**	.351***	.412****	.070
IRSSCB	.200	.109	.150	.206	-.025
E	-.120	-.062	-.044	.082	.071
N	-.034	.021	.010	-.106	-.063
L	-.053	-.178	-.095	-.027	.216
P	.106	.109	.102	.133	.038
So	.003	.002	-.160	-.067	-.001

**** $p < .001$

*** $p < .01$

** $p < .02$

* $p < .05$

All calculations were performed using SPSS. As well as the sign and magnitude of each correlation, the tables present the results of two-tailed tests of the significance of the coefficients from zero, since in most cases directional hypotheses were not formulated. In the interests of completeness, the particular probability value associated with each correlation is reported but in the discussion which follows the .05 rejection region is employed. The tables are considered with reference to the hypotheses guiding the study.

Hypothesis 1 (in the null form): There is no relationship between indices of reactivity and the likelihood of being detected; more specifically, the level and amplitude of SC are unrelated to Differential Responsiveness or Detectability under any of the conditions examined. An intuitive analysis of the problem suggested as the research hypothesis that reactivity and the indices of Detectability are positively correlated. On the other hand an analysis based on the psychometrics of difference scores suggested that a negative correlation was possible.

Inspection of the three tables indicates that the null hypothesis must be rejected for the level index (SCBL) in the following cases:

(a) in the card test, for the mute condition, where Detectability is assessed for control subjects (DETCA3);

(b) in the mock agent procedure, for the GKT in both no and mute conditions where Differential Responsiveness is assessed (DIFRA4 and DIFRA5);

(c) in the mock agent procedure, for the GKT in the mute condition where Detectability is assessed (DETCA5).

They cannot be rejected in other instances. However, mention should be made of the correlations with Detectability for yes and no conditions for control subjects. Although not statistically significant, the magnitude and sign of these coefficients are consistent with that for control subjects under the mute condition.

For the amplitude index (SCAMP) the null hypothesis must be rejected:

(a) in the card test for the mute condition, where both Differential Responsiveness and Detectability are assessed for control subjects (DIFRA3 and DETCA3);

(b) in the mock agent procedure, for the GKT in the no condition for both measures of Detectability (DIFRA4 and DETCA4);

(c) in the mock agent procedure for the CQ test (DRC12).

The null hypothesis cannot be rejected in all other cases.

The pattern of results is consistent with the first research hypothesis (a positive relationship) in most cases in the mock agent procedure, but not in the card test. Only in the case of control subjects in the mute condition is there any support. The conclusion to be drawn is that, in general, reactivity is positively related to the likelihood of being detected in the mock agent procedure. There is also some suggestion that control subjects who are reactive are likely to be incorrectly detected in the card test.

Hypothesis 2 (in the null form): There is no relationship between indices of lability and the likelihood of being detected; more specifically, the frequency of non-specific and evoked responses are unrelated to Differential Responsiveness or Detectability under any of the conditions examined. The intuitive analysis of the problem suggested as the research hypothesis that the lability indices and the indices of Detectability are positively correlated. On the other hand, an analysis in terms of OR theory suggested a negative correlation.

Inspection of the tables of correlations indicates that for the index of non-specific responses (NSR), the null hypothesis must be rejected in the following cases:

(a) in the card test, for experimental subjects in the no and mute conditions for Differential Responsiveness (DIFRA2 and DIFRA3);

(b) in the card test, for experimental subjects in the mute condition for Detectability (DETCA3);

(c) in the mock crime, for the GKT in the no condition for Differential Responsiveness (DIFRA4);

(d) in the mock crime, for GKT in the mute condition for Detectability (DETCA5).

For the index of frequency of evoked responses (TNR), the null hypothesis must be rejected in the following cases:

(a) in the card test, for experimental subjects in the no condition for Differential Responsiveness (DIFRA2);
(b) in the card test, for experimental subjects in the yes and mute conditions, for Detectability (DETCA1 and DETCA3);
(c) in the mock crime, for the GKT, for Differential Responsiveness and Detectability in both no and mute conditions (DIFRA4, DIFRA5, DETCA4, and DETCA5).

The conclusion to be drawn is that lability is positively related to the likelihood of detection in the mock agent procedure when the GKT is used. Less clearly, lability is positively related to the likelihood of detection in the card test for experimental subjects.

Hypothesis 3 (in the null form): There is no relationship between the relative reactivity (specificity) of the system and the likelihood of being detected; more specifically, subjects who are relatively more reactive in SC do not show greater Differential Responsiveness or Detectability under any of the conditions examined. The intuititive analysis of the problem suggested as the research hypothesis a positive correlation between specificity and Detectability.

Inspection of the three tables indicates that for the specificity index (IRSSCB) the null hypothesis must be rejected in the following cases:

(a) in the card test, for experimental subjects, for Differential Responsiveness in all conditions (DIFRA1, DIFRA2, and DIFRA3);
(b) in the card test, for experimental subjects, for Detectability only in the mute condition (DETCA3);
(c) in the card test, for control subjects, for Detectability in the yes condition (DETCA1).

The conclusion to be reached is that specificity is positively related to the likelihood of experimental subjects being detected in the card test on the Differential Response index. There is no relation between the likelihood of being detected on a GKT or a CQ test and IRSSCB. The poor relation demonstrated between the Detectability index (DETCA) and IRS in the card test permits no firm conclusion.

Hypothesis 4 (in the null form): There is no relationship between extraversion and the likelihood of being detected; more specifically, score on the E scale of the EPQ is unrelated to Differential Responsiveness or Detectability under any of the conditions examined. Consideration of previous research suggested as a research hypothesis that E scale score and the likelihood of detection are negatively correlated.

Examination of the three tables indicates that the null hypothesis cannot be rejected for any index in any condition of the card test or mock agent paradigm.

Hypothesis 5 (in the null form): There is no relationship between socialization and the likelihood of being detected; more specifically, score on the socialization (So) scale of the CPI is unrelated to Differential Responsiveness or Detectability under any of the conditions examined. Consideration of previous research suggested that So scale score and Detectability are negatively correlated.

The null hypothesis cannot be rejected for any index, in any condition of the card test or mock agent paradigm. It should be noted, however, that in the card test, for the no condition, where Detectability is assessed for experimental subjects (DETCA2) the null hypothesis comes close to rejection.

Although not predicted, statistically significant correlations with score on the N scale of the EPQ were observed under the following conditions:

(a) in the card test, for the yes and mute conditions where Differential Responsiveness is assessed for control subjects (DIFRA1 and DIFRA3);
(b) in the card test, for the mute condition where Detectability is assessed for control subjects (DETCA3).

As the correlations in (a) varied in sign, little can be made of this.

Multivariate Analyses (Electrodermal Responsiveness)

The primary purpose of the multivariate analyses was to determine the extent to which each of the sets of indices of electrodermal activity (reactivity, lability, and specificity) can be considered to contribute uniquely to variance in each of the measures of Differential Responsiveness and Detectability. Although the bivariate analyses just considered indicated a number of significant relationships for indices from each set it could be that some of these are redundant. Lability and reactivity indices while separating out in the factor analysis were not completely independent of each other. It is possible therefore that a more limited set of correlations than those found in the bivariate analyses can account for the relationships with the various indices of deception. To test this possibility a series of multiple regression analyses were performed in which the indices of Differential Responsiveness and Detectability served as criteria and the measures of reactivity, lability, and specificity as predictors. Three models or predictor sets were constructed for testing. The first included only the two reactivity measures as predictors, the second added the two lability indices to the two reactivity indices and the third added the index of specificity to the predictor set. Comparison of the predictive power of these models allowed determination of the extent to which lability and specificity added to that afforded by reactivity. The ordering of models for testing, although arbitary, was based on the premise that the simplest explanation of individual differences in the likelihood of detection is that due to individual differences in the absolute reactivity of the system. The addition of lability and specificity makes for a less parsimonious account and must therefore be justified on the grounds that inclusion of these predictors significantly increases the total amount of variance in the criterion which is predictable.

The analyses are summarized in Tables 6.4 to 6.20. Each table presents for each model the multiple correlation (R), the proportion of variance (PV) in the criterion accounted for by the model, and the *F* value and degrees of freedom (*df*) for the test of R against zero. These analyses were performed using the stepwise regression procedure in SPSS. The tables also present for the four- and five- predictor models the increment in the proportion of variance (Increment in PV) in the criterion accounted for by inclusion of the extra predictors in the model, and the *F* value and *df* for the significance of the increment. The increment in PV was obtained by subtraction and the *F* test was calculated using the formula of Cohen and Cohen (1975, p.136). No correction (shrinkage) of the multiple Rs was applied, as the purpose of the analysis was not to provide a prediction battery for use in subsequent studies. Such an exercise, if considered worthwhile, would require standardization and cross validation beyond the scope of the present study. The uncorrected Rs are therefore reported as summaries of the relationships obtained. It should be noted in reviewing these analyses that the reliabilities of the criterion variables are low (cf. Tables 5.13 and 5.27) and as a consequence substantial predictive accuracy cannot be expected.

Tables 6.4 to 6.9 present summaries of the analyses for the experimental subjects in the card test. The best results were obtained in the mute condition. Inspection of Table 6.6 indicates that for experimental subjects 27% of the variance in Differential Responsiveness in the mute condition of the card test (DIFRA3) was accounted for by the five predictor variables. Further, the multiple R for this model, as well as for the two and four predictor models, was statistically significant. The addition of the lability indices results in the total proportion of variance accounted for (PV) increasing by more than two and a half times that accounted for by the reactivity indices. The increment in PV (15%) is statistically significant. The lability indices are thus contributing to the prediction of Differential Responsiveness in the mute condition, over and above that afforded by the reactivity indices. The addition of the specificity variable makes no substantial or significant difference to PV, the increment being only some 4%. Thus the finding in the bivariate analysis of a statistically significant correlation (0.441) between the specificity index and DIFRA3 must, in the light of this analysis, be considered redundant with the statistically significant correlation between lability and the criterion.

Inspection of Table 6.9 leads to a similar conclusion in the case of the Detectability index in the mute condition of the card test (DETAC3). Here, the addition of the lability indices increases the PV by almost three times that accounted for by the reactivity indices. The increment in PV (16%) is again substantial and significant. Once again the addition of the specificity variable makes no significant or substantial difference to PV. The finding in the bivariate analysis of a statistically significant correlation between the specificity index and this criterion (DETCA3) is thus also redundant with the statistically significant correlation with lability.

Inspection of Table 6.5 indicates that in the case of Differential Responsiveness in the no condition of the card test the two variable model failed to produce a statistically significant multiple R. However, the addition of the lability indices more than doubles PV and gives rise to a statistically significant multiple R. The increment in PV (8%) although substantial is not statistically significant. The addition of the specificity variable makes a further increase in PV though again the increment is not significant. The conclusions to be drawn from this condition must be more equivocal than in the mute condition, as the increment in PV for the lability indices is not statistically significant, and the increment for the specificity variable is of the same order as that for lability. The trend of the results is broadly consistent with that found in the mute condition, but the conclusions cannot be drawn with the same conviction.

This is not the case with the Detectability index in the no condition. Here the same conclusions cannot be drawn. As an inspection of Table 6.8 shows, none of the models produce a statistically significant multiple R. Nor do the lability indices add to PV in other than a trivial way.

The same can be said for prediction of Differential Responsiveness in the yes condition (see Table 6.4). However, in this case there is some suggestion that the specificity variable may be contributing to the prediction. PV more than doubles when the specificity index is added to the predictive equation, but the increment in PV is not statistically significant. When the criterion is Detectability in the yes condition (Table 6.7) the prediction is uniformly poor. The addition of the lability indices leads to some improvement in prediction but the increment is not statistically significant.

The results for control subjects in the card test appear in Tables 6.10 to 6.15. In the case of control subjects, prediction was again best in the mute condition, but here the major predictor set was the reactivity indices and little was contributed by lability. For both criteria of Differential Responsiveness and Detectability (Tables 6.12 and 6.15) approximately 18% of the variance is accounted for by the two predictor model. The multiple R in both cases is statistically significant, and, although the Rs in the case of the four- and five-variable models are also significant, in no case are the increments

Table 6.4

Summary of Step-Wise Multiple Regression Analysis using DIFRA1 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.1239	1.54	.56	2/72			
2 SCBL + SCAMP + NSR + TNR	.2115	4.47	.82	4/70	2.93	.52	4/68
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.3370	11.36	1.77	5/69	6.89	1.01	5/65

Table 6.5

Summary of Step-Wise Multiple Regression Analysis using DIFRA2 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.2819	7.95	3.11	2/72			
2 SCBL + SCAMP + NSR + TNR	.4042	16.34	3.42^{*}	4/70	8.39	1.71	4/68
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.4762	22.68	4.05^{**}	5/69	6.34	1.07	5/65

** $p < .01$

* $p < .05$

Table 6.6

Summary of Step-Wise Multiple Regression Analysis

using DIFRA3 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.2924	8.55	3.37*	2/72			
2 SCBL + SCAMP + NSR + TNR	.4878	23.80	5.46**	4/70	15.25	3.40*	4/68
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.5233	27.39	5.20**	5/69	3.59	.64	5/65

** $p < .01$

* $p < .05$

Table 6.7

Summary of Step-Wise Multiple Regression Analysis

using DETCA1 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.1246	1.55	.57	2/72			
2 SCBL + SCAMP + NSR + TNR	.2867	8.22	1.57	4/70	6.67	1.24	4/68
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.3163	10.00	1.53	5/69	1.78	.26	5/65

Table 6.8

Summary of Step-Wise Multiple Regression Analysis using DETCA2 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.1306	1.71	.62	2/72			
2 SCBL + SCAMP + NSR + TNR	.2062	4.25	.78	4/70	2.54	.45	4/68
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.2616	6.85	1.01	5/69	2.60	.36	5/65

Table 6.9

Summary of Step-Wise Multiple Regression Analysis using DETCA3 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.2893	8.37	3.29^{*}	2/72			
2 SCBL + SCAMP + NSR + TNR	.4928	24.28	5.61^{**}	4/70	15.91	3.59^{*}	4/68
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.4928	24.29	4.43^{**}	5/69	.01	0	5/65

** $p < .01$

* $p < .05$

in PV with addition of the lability or specificity variables either sizable or significant. Hence for both criteria in the mute condition the reactivity indices are the major predictors. This is perhaps not surprising in view of the findings in the bivariate analyses of significant correlations between certain of the reactivity indices and the criteria but of no significant relationships with the lability indices.

Inspection of Tables 6.10 and 6.13 indicates that in the yes condition none of the predictors in the two-, four-, or five-variable models significantly correlate with the Differential Response or the Detectability criterion. The correlation revealed in the bivariate analysis between the specificity variable and Detectability in the yes condition for control subjects must, in the light of Table 6.13, be viewed with some caution. The increment in PV which the specificity index provides over and above that provided once the lability indices are included is slight.

In the case of Differential Responsiveness and Detectability in the no condition (Tables 6.11 and 6.14), none of the predictors in the two-, four-, or five-predictor models significantly correlate with either criterion. There were no statistically significant correlations between predictors and these two criteria in the bivariate analysis.

The findings of these multiple regression analyses can be summarized as follows:

(1) Predictability is best in the mute condition for both experimental and control subjects.

(2) For experimental subjects lability measures add significantly and substantially to predictability on the basis of reactivity alone.

(3) This is not true for control subjects for whom reactivity is the major predictor.

(4) There is little evidence that specificity adds other than a trivial proportion to the variance predictable from reactivity and lability.

The regression analyses for the data from the mock agent procedure are summarized in Tables 6.16 to 6.20. Inspection of Table 6.16 indicates that some 37% of the variance in Differential Responsiveness in the no condition of the GKT was accounted for by the five-predictor

Table 6.10

Summary of Step-Wise Multiple Regression Analysis using DIFRA1 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.1806	3.26	.67	2/40			
2 SCBL + SCAMP + NSR + TNR	.2960	8.76	.91	4/38	5.50	.55	4/36
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.2967	8.80	.71	5/37	.04	0	5/33

Table 6.11

Summary of Step-Wise Multiple Regression Analysis using DIFRA2 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.2214	4.90	1.03	2/40			
2 SCBL + SCAMP + NSR + TNR	.2775	7.70	.79	4/38	2.80	.27	4/36
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.3221	10.37	.86	5/37	2.67	.19	5/33

Table 6.12

Summary of Step-Wise Multiple Regression Analysis using DIFRA3 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.4193	17.58	4.27*	2/40			
2 SCBL + SCAMP + NSR + TNR	.4757	22.63	2.78*	4/38	5.05	.59	4/36
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.5620	31.59	3.42*	5/37	8.96	.87	5/33

* $p < .05$

Table 6.13

Summary of Step-Wise Multiple Regression Analysis using DETCA1 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.3562	12.69	2.91	2/40			
2 SCBL + SCAMP + NSR + TNR	.4291	18.41	2.14	4/38	5.72	.63	4/36
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.4574	20.93	1.96	5/37	2.52	.21	5/33

Table 6.14

Summary of Step-Wise Multiple Regression Analysis using DETCA2 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.2742	7.52	1.63	2/40			
2 SCBL + SCAMP + NSR + TNR	.3780	14.29	1.58	4/38	6.77	.71	4/36
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.3831	14.67	1.27	5/37	.38	.03	5/33

Table 6.15

Summary of Step-Wise Multiple Regression Analysis using DETCA3 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.4290	18.40	4.51*	2/40			
2 SCBL + SCAMP + NSR + TNR	.4836	23.39	2.90*	4/38	4.99	.59	4/36
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.5050	25.51	2.53*	5/37	2.12	.19	5/33

* $p < .05$

model, and that the multiple R in this case was statistically significant. The two-variable and four-variable models were also statistically significant. Examination of the increment in variance accounted for indicates that addition of the lability indices results in a 9% increase in PV for the Differential Responsiveness criterion. Although not statistically significant this increase is less than trivial and suggests that the lability indices are contributing to the prediction of Differential Responsiveness over and above the reactivity indices. The addition of the specificity variable makes no significant or substantial difference. A similar conclusion is reached in the case of the Detectability index in the no condition of the GKT (see Table 6.18). The evidence for a unique contribution from the lability indices is strengthened in this case by the significant F for the increment of the four variable over the two variable model.

In the case of Detectability in the mute condition (Table 6.18), the lability indices again appear to make a substantial though not statistically significant contribution to PV. The four variable model more than doubles PV and gives rise to a statistically significant multiple R which is not obtained with the two-variable model. The same cannot be said for the lability indices in the case of Differential Responsiveness in the mute condition (Table 6.17). The difficulty here, however is that none of the predictors in the two-, four-, or five-variable combinations correlate significantly with the criterion. This does not of course contradict the univariate findings of correlations for SCBL and TNR since these were obtained for the variables taken singly rather than in combination.

Multiple prediction was also disappointing in the case of the CQ test (Table 6.20). None of the models gave rise to a statistically significant multiple R, and hence conclusions about the relative importance of predictors cannot be reached. Again, this does not make the univariate finding of a statistically significant correlation between SCAMP and DRC12 redundant.

Table 6.16

Summary of Step-Wise Multiple Regression

Analysis using DIFRA4 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.5169	26.72	12.76**	2/70			
2 SCBL + SCAMP + NSR + TNR	.5986	35.83	9.49**	4/68	9.11	2.35	4/66
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.6057	36.68	7.76**	5/67	.85	.17	5/63

** $p < .01$

Table 6.17

Summary of Step-Wise Multiple Regression

Analysis using DIFRA5 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.2597	6.74	2.53	2/70			
2 SCBL + SCAMP + NSR + TNR	.2899	8.41	1.56	4/68	1.67	.30	4/66
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.2910	8.47	1.24	5/67	.06	.01	5/63

Table 6.18

Summary of Step-Wise Multiple Regression

Analysis using DETCA4 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.2520	6.35	2.37	2/70			
2 SCBL + SCAMP + NSR + TNR	.4483	20.09	4.28**	4/68	13.74	2.84*	4/66
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.4565	20.84	3.53**	5/67	.75	.12	5/63

** p < .01

* p < .05

Table 6.19

Summary of Step-Wise Multiple Regression

Analysis using DETCA5 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.2658	7.07	2.66	2/70			
2 SCBL + SCAMP + NSR + TNR	.4231	17.91	3.71**	4/68	10.84	2.19	4/66
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.4438	19.70	3.29**	5/67	1.79	.28	5/63

** p < .01

Table 6.20

Summary of Step-Wise Multiple Regression Analysis using DRC12 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
1 SCBL + SCAMP	.2704	7.31	2.76	2/70			
2 SCBL + SCAMP + NSR + TNR	.3228	10.42	1.98	4/68	3.11	.57	4/66
3 SCBL + SCAMP + NSR + TNR + IRSSCB	.3268	10.68	1.60	5/67	.24	.04	5/63

The results of these analyses can be summarized as follows:

(1) Predictability was best for the no condition of the GKT

(2) Lability adds substantially to the prediction afforded by reactivity alone in the GKT

(3) Detectability in the CQ test is virtually unpredictable using the predictor set employed here.

Multivariate Analyses (Personality Scores)

The purpose of these analyses was to assess the extent to which the self report measures added to the prediction of Detectability afforded by the indices of electrodermal activity. The analyses reported in Chapter 4 pointed to little overlap between the indices of responsiveness and the self report variables. The exceptions were that high scorers on N, P, and socialization showed somewhat greater reactivity. Given this degree of independence, it is possible that the self report tests might add to the prediction of Detectability even though the bivariate correlations with the criterion were for the most part quite low. Such a result would imply that personality exerts an influence through factors independent of those to do with the responsiveness of the system, such as fear of the consequences of detection or confidence in the procedures employed (cf Lykken, 1974). The results of these analyses were, however, almost wholly negative as inspection of summary tables 6.21 to 6.37 indicates.

Since the samples for which personality data were available were smaller than those for which data on the responsiveness indices were available, the regression analyses were all re-calculated using the smaller _n_. This resulted in some variation in the R's obtained when the electrodermal indices were used as the predictor set compared to those reported above, a variation attributable to sampling error. All comparisons of the relative predictive power of the personality measures with that of the electrodermal measures were based on these new regression analyses.

The new analyses specified the first three steps employed in the earlier analyses and two further steps: the addition of the four EPQ variables, and the addition of the CPI socialization score. This

Table 6.21

Summary of Step-Wise Multiple Regression Analysis using DIFRA1 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.4264	18.18	.72	9/29	3.98	.13	9/24
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.4274	18.26	.63	10/28	.08	0	10/19

Table 6.22

Summary of Step-Wise Multiple Regression Analysis using DIFRA2 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.5788	33.51	1.62	9/29	8.01	.32	9/24
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.5911	34.94	1.50	10/28	1.43	.04	10/19

Table 6.23

Summary of Step-Wise Multiple Regression Analysis using DIFRA3 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.6098	37.19	1.91	9/29	9.82	.42	9/24
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.6173	38.10	1.72	10/28	.91	.03	10/19

Table 6.24

Summary of Step-Wise Multiple Regression Analysis using DETCA1 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.4383	19.21	.77	9/29	6.73	.22	9/24
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.4383	19.21	.67	10/28	0	0	10/19

Table 6.25

Summary of Step-Wise Multiple Regression Analysis using DETCA2 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.4620	21.34	.87	9/29	7.50	.25	9/24
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.5238	27.44	1.06	10/28	6.10	.16	10/19

Table 6.26

Summary of Step-Wise Multiple Regression Analysis using DETCA3 as the Criterion for Experimental Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.6608	43.66	2.50^{*}	9/29	5.48	.26	9/24
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.6629	43.94	2.20^{*}	10/28	.28	.01	10/19

* $p < .05$

Table 6.27

Summary of Step-Wise Multiple Regression Analysis using DIFRA1 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.5114	26.15	1.30	9/33	17.35	.73	9/28
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.5227	27.32	1.20	10/32	1.17	.04	10/32

Table 6.28

Summary of Step-Wise Multiple Regression Analysis using DIFRA2 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.3688	13.60	.58	9/33	3.23	.12	9/28
5 SCBL +SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.3883	15.08	.57	10/32	1.48	.04	10/23

Table 6.29

Summary of Step-Wise Multiple Regression Analysis using DIFRA3 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.6984	48.77	3.49^{**}	9/33	17.18	1.04	9/28
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.7028	49.39	3.12^{**}	10/32	.62	.03	10/23

** p < .01

Table 6.30

Summary of Step-Wise Multiple Regression Analysis using DETCA1 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.4947	24.47	1.19	9/33	3.54	.14	9/28
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.5244	27.49	1.21	10/32	3.02	.10	10/23

Table 6.31

Summary of Step-Wise Multiple Regression Analysis using DETCA2 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.4271	18.25	.82	9/33	3.58	.14	9/28
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.4323	18.69	.74	10/32	.44	0	10/23

Table 6.32

Summary of Step-Wise Multiple Regression Analysis using DETCA3 as the Criterion for Control Subjects

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.5934	35.21	1.99	9/33	9.70	.47	9/28
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.6108	37.31	1.90	10/32	2.10	.08	10/23

Table 6.33

Summary of Step-Wise Multiple Regression Analysis using DIFRA4 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.6329	40.06	4.68**	9/63	3.38	.37	9/58
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.6541	42.79	4.64**	10/62	2.73	.25	10/53

** $p < .01$

Table 6.34

Summary of Step-Wise Multiple Regression Analysis using DIFRA5 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.3713	13.79	1.12	9/63	5.32	.40	9/58
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.3835	14.71	1.07	10/62	.92	.06	10/53

Table 6.35

Summary of Step-Wise Multiple Regression Analysis using DETCA4 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.4726	22.33	2.01	9/63	1.49	.13	9/50
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.4726	22.34	1.78	10/62	.01	0	10/53

Table 6.36

Summary of Step-Wise Multiple Regression Analysis using DETCA5 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.4673	21.84	1.96	9/63	2.14	.18	9/58
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.4682	21.92	1.74	10/62	.08	.01	10/53

ordering was chosen on the assumption that the EPQ variables provide a more general statement of personality factors with-in which the socialization measure can be located. The correlations of N and P with socialization, revealed in Chapter 4, support this view. Tables 6.21 to 6.37 are limited to the results for steps 4 and 5 of each analysis.

Inspection of these tables indicates that the greatest contribution of the EPQ variables was made in the case of the Differential Responsiveness indices for control subjects in the yes and mute conditions of the card test. The increment in neither case was statistically significant. For socialization, the analyses indicated that only a trivial proportion of the variance in Detectability was unique to this measure once the EPQ variables had been removed.

Table 6.37

Summary of Step-Wise Multiple Regression

Analysis using DRC12 as the Criterion

Step	R	PV	F	df	Increment in PV	F	df
4 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L	.3946	15.57	1.29	9/63	4.89	.37	9/58
5 SCBL + SCAMP + NSR + TNR + IRSSCB + E + N + P + L + So	.3946	15.57	1.14	10/62	0	0	10/53

Discussion

None of the measures of individual differences in electrodermal responsiveness showed a totally consistent pattern of intercorrelation with the criterion measures. Both the type of criterion measure, Differential Responsiveness or Detectability, and the conditions under which it was obtained influenced the correlation. While some of the fluctuation in the magnitudes of the correlation coefficients are no doubt due to errors of measurement in that the criteria evidenced a less than satisfactory degree of reliability, this is unlikely to account for all the fluctuations. For example, the number of significant correlations with the Differential Responsiveness index was not clearly greater than that with the Detectability index, and yet the former would be expected to be more reliable than the latter, an expectation supported by the estimates reliability reported in Tables 5.13 and 5.27. The results in general thus point to a greater complexity than provided for in the original analysis of the problem.

One of the factors which appears to moderate relationships, at least for the card test data, is whether or not the subject is required to respond to the stimulus. Where no response was required, the mute condition, the measures of electrodermal activity accounted for more of the variance in the criteria. This was true for both criterion measures and for experimental and control subjects. This did not hold for the mock agent procedure, where greater predictability was found under the no and not the mute condition of the GKT. There was, however, an unfortunate confounding of instructional set with order of presentation in the mock agent procedure, which prevents a direct comparison with the findings of the card test on this point. Order of presentation was found in Chapter 5 to exert a strong influence on responsiveness.

One possible explanation for the greater predictability in the mute condition is that where no verbal response is required of the subject, individual differences factors exert a greater influence. Where all subjects are required to lie to the examiner the influence that this has on subjects' state, presumably an increase in arousal, serves to reduce differences among subjects which are important under more neutral conditions. Such an interpretation is consistent with the view that trait measures are most important where situational factors do not constrain the nature of the response to be made (Stagner, 1977).

This may also be suggested as a reason for the failure of the responsiveness measures to predict detectability in the CQ test. Order of presentation and any habituation effect that this might induce cannot be invoked to explain the poor predictability here since the CQ test was presented first in the interrogation. The direct lie which is required in the CQ test might, however, constrain the range of variability among subjects and hence lessen the importance of individual difference factors. Speculation of this sort is perhaps premature until such time as the moderating effect of conditions observed in the present study is replicated.

Of the predictions examined, lability was clearly the most useful and the personality tests the least useful. The result with labilty is consistent with that reported by Waid and colleagues (Waid & Orne, 1980; Waid, Wilson, & Orne, 1981). In the chapter that follows an attempt is made to give this result some basis in theory. For the present, it is necessary simply to note that the contribution of lability is largely independent of that of reactivity and is confined to the predicton of Differential Responsiveness where detection is in fact involved (i.e., for experimental and not control subjects). This discrimination increases confidence that lability is a genuine correlate of the likelihood of detection.

Finally, the lack of efficiency of the self report tests as predictors should be noted. Neither the results of Waid, Orne, and Wilson with the socialization scale nor those of Balloun and Holmes with the EPI E scale were replicated here. Little previous work has been reported on the correlations between these measures and indices of PDD, but if the literature on personality and psychophysiological responsiveness in general is any guide the yield of non-significant findings in this study is not likely to be atypical.

CHAPTER 7

CONCLUSIONS

The first task of the research programme was the establishment of dimensionality in the electrodermal system. Podlesny and Raskin (1977) had identified reactivity, lability, and specificity as major dimensions of responsiveness of consequence for PDD, but little systematic investigation of this issue had been reported in the literature. The work of Martin and Rust (1976) and Lockhart and Lieberman (1979) pointed to a distinction between magnitude and level measures on the one hand and measures of rate of change and nonspecific activity on the other. The distinction was not however a strong one as the comment of Martin and Rust regarding the importance of a general factor of reactivity made clear. Moreover, the question of independence of specificity from the other two had nowhere been addressed. The correlational analyses reported in Chapter 4 sought to clarify these issues.

The results of the correlational analyses indicate that a lability factor can be separated from a general factor of reactivity, but that the further separation of a specificity factor is not possible. The measure of specificity employed was found to correlate with both measures of reactivity and lability. The former relationship is not perhaps surprising as both involve consideration of response level, one in an absolute and the other in a relative sense. The relationship with the latter is more interesting as it suggests that a component of lability involves the prominence of the electrodermal system vis a vis other autonomic response systems in the individual's reaction to or interaction with the environment. Whether this reflects a constitutional basis or arises as a result of the learning history of the individual is a matter about which speculation is best reserved until the relationship is replicated elsewhere. For the present, the correlational analysis simply asserts that a two-factor rather than a three-factor model of electrodermal responsiveness is a better fit to the available data.

As discussed in Chapter 2, a two-factor model of the sort revealed in the correlational analysis can be considered an analogue in differential psychophysiology of the two-process model of orienting behaviour and attention discussed in the wider psychophysiological literature. One factor is the counterpart of the general arousal process and the other of the stimulus processor. The relationship can be clearly seen in the case of Sokolov's classic model in which a stimulus comparator and arousal system are linked in a negative feedback arrangement to modulate response output. The use of Sokolov's model as the exemplar makes relatively easy the identification of lability with individual differences in stimulus processing and reactivity with individual differences in arousal, although it must be acknowledged that such an identification is conceptually and not empirically based. The reactivity factor is linked with operation of the arousal system because of the prevailing view in the literature of the relevance of at least one of the defining measures of the reactivity factor, response level, as an operational index of arousal. The linking of lability to the stimulus processor aspect of two-process models such as Sokolov's is more contentious, since one of the defining measures of this factor, frequency of NSRs, has been discussed as an index of arousal along with response level. It is linked here with the stimulus processor because speed of habituation, the other defining index of lability, depends critically on stimulus analysis. Although habituation involves the operation of both analysis and arousal systems, the variance left in an habituation measure when the arousal component has been removed must, in terms of the theory, depend on stimulus analysis. The same can be said for frequency of NSRs. Although dependent on the general level of arousal, variance in the NSR index which remains after an arousal factor has been extracted can be attributed to stimulus analysis.

An adequate theoretical description of the factors of electrodermal responsiveness is necessary, of course, if the relationships of these measures with indices of Detectability (or the lack of relationship) is to be understood. The description of reactivity in terms of individual differences in the arousal process or background level or state of the electrodermal system needs little elaboration. The description of lability in terms of stimulus analysis, however, calls for some comment as the question immediately arises: Individual differences in what type of analysis are being reflected in the lability measures? The answer provided by Sokolov's earlier statement of his theory is stimulus novelty, but this position has been found wanting by more recent theorists who argue that it is stimulus significance which is the essential feature to which analysis is directed (see O'Gorman, 1977 for a review of this issue). It is the view of Maltzman (1979), for example, that individual differences in frequency of NSRs reflect differences in the significance which individuals attribute to the surroundings in which the measure is obtained, with labiles attributing greater significance than stables. In the same way, slow habituation of response is seen by Maltzman to result from the subject continuing to consider the stimulus for habituation significant in some way. Such a view would seem particularly relevant to the study of PDD since novelty is rarely a feature of the stimulus material employed in tests of PDD whereas the significance of the material is crucial.

The conclusion to be drawn from this analysis and the empirical work reported in Chapter 4 is that there are two basic dimensions of electrodermal responsiveness, termed reactivity and lability, which reflect individual differences in arousal level, at least within the electrodermal system (a qualification made necessary by the observation that autonomic measures seldom correlate), and individual differences in processing the significance of stimulus events. The clear implication is that only these two dimensions need be referred to in accounts of individual differences in PDD which involve tests of Differential Responsiveness within the electrodermal system. The potentially troublesome problem of specificity can thus be subsumed within the two dimensional model.

The second task of the research programme was to develop measures of Detectability which could be used as criteria for prediction from the dimensions of electrodermal responsiveness. Two sets of measures were derived using data both from a card test and a mock agent test. One set of measures (prefixed DETC) was derived using the procedure followed in blind assessment of a PDD record, i.e., determination of the response of largest magnitude. The second (prefixed DIFR) employed as much of the information available as possible by averaging over the non-critical items. The latter, not surprisingly, proved a somewhat "better" set of measures, although because these measures cannot be derived by an examiner blind with respect to the critical stimuli they cannot be readily generalized to the field situation.

These measures were acceptable criteria for a study of the correlates of individual differences in PDD in as much as the accuracy rates for detection were broadly comparable with those reported in the literature. The rates obtained were in general lower than those reported in published studies but the discrepancies were not so great as to question seriously the representativeness of the measures.

On a second point, however, the acceptability of the measures as criteria can be criticized. In no case did a measure show the degree of reliability expected of a measure of individual differences by psychometricians. Reliability is of course an important characteristic of any criterion since low reliability attenuates the correlations with other variables to be used as predictors. Reliability in the test-retest sense has not been reported on previously and it is not possible therefore to establish whether the present findings are typical of measures of PDD. Where reliability has been discussed previously it has been in the inter-judge sense, i.e., where two or more assessors score the same record. Reliability in this sense has been found to be high, often as high as 100% (Barland & Raskin, 1975; Podlesny & Raskin, 1978). The two types of reliabilities are, however, independent as one (test-retest) concerns sources of measurement error arising within the subject and the other (inter-rater) concerns sources within the judge.

In the present study reliability was assessed by correlating response measures across two trials. The trials were separated temporally by only a few minutes. Certainly it would be more convincing to demonstrate reliability over a longer time interval, such as a week, but reliability over a short interval would seem to be a prerequisite for reliability over a longer interval. It could be argued of course that the short time period between trials may not be the optimal method for demonstrating reliability, since response habituation has to be contended with over short intervals. Subjects may be more difficult to detect on trial 2 because habituation has substantially lowered response amplitude. Such a loss in sensitivity of the measure, it could be argued, is responsible for its apparently low reliability. For this argument to hold, however, the difference in response between target and non-target cards should be lower on trial 2 than on trial 1. That is, the argument implies not only that response amplitude habituates absolutely but also that Differential Responsiveness habituates. An analysis of the present data shows that this is not the case. As inspection of Table 5.4 indicates there is little difference between trials. Responsiveness decreases but the differential is maintained.

The second and more likely explanation for the low reliability coefficients involves the very limited sample of behaviour on which the estimates of reliability are based. Psychometrics (as well as common sense) suggests that the larger the sample of behaviour the more reliable is the estimate based on it. Psychometric tests, therefore, include a large number of items. The tests of deception used here, and these are not atypical, include few items (questions) on which to base a decision of deceptiveness. For example, the card test involves the asking of one critical question per trial. The problem of limited sampling is compounded by the use of a difference score as the basic measure in determining deceptiveness. As noted in Chapter 2, difference scores are notoriously unreliable because they add the unreliability of each of the components. A systematic study of reliability should pay particular attention to the conditions under which detectability is assessed. One criticism that could be leveled at this study is that not enough emphasis was attached to the control questions. Greater emphasis attached to the control questions might have altered accuracy rates and the reliability of detection.

This analysis suggests that the unreliability found here is not surprising. The clear implication is that a systematic study of the reliability of PDD measures and of ways of improving it should be undertaken. Such a study should examine as a first priority the influence of test length on reliability since it may be that arguments from test theory do not hold when applied to physiological response systems. Such a study would answer the question of the generality of detectability as a trait. It may well be that there is no such generalized trait and that the detection of deception is an episodic effect determined by the state of the subject at the time of the test and even from moment to moment within the test. The results of the present study suggest that such a high degree of specificity is unlikely, but the limit of this should be determined empirically.

The third task of the research programme was the bringing together of the measures of electrodermal responsiveness and the criteria set. This exercise in prediction was guided by the few empirical studies available in the literature and the little that could be gleaned from current theorizing about PDD. These two sources of hypotheses were not always in agreement as the discussion in Chapter 2 sought to point out. A major discrepancy occurred in the case of lability, which the work of Waid, Orne, and colleagues indicated should correlate positively with detectability whereas the implication from Ben-Shakhar's theory was the reverse. The results outlined in Chapter 6 provide support for previous results rather than for the theoretical deductions.

Before examining this apparent conflict of theory and empirical result the point should be made that the findings of the present programme do extend those currently available in the literature. At the time of commencing this research programme the work of Waid, Orne, and colleagues on lability as a correlate of detectability was not available. It is significant therefore that results consistent with theirs were obtained as these provide independent corroboration of the importance of lability. More significantly, they demonstrate that the contribution of lability is independent of reactivity or arousal which might be thought of as a more basic parameter of electrodermal responsiveness. By partialling reactivity from lability in the multiple regression analyses, which had not previously been undertaken, the "true" contribution of lability per se could be assessed. Although the correlations obtained were not high, a fact to be accounted for at least partly by the low reliability of the criterion measures, they were obtained with sufficient frequency to indicate that the correlation between lability and detectability is reasonably robust. Moreover, the data from the card test point to the relevance of lability only when subjects are actually attempting to conceal information. The correlation is thus specific to Detectability and not a feature of Differential Responsiveness however assessed. Reactivity, it should be noted, did not demonstrate such specificity.

It should also be noted that reactivity and Detectability were positively correlated though the correlations were seldom significant. Although intuitively a positive correlation between the two is to be expected, it was argued in Chapter 2 that in terms of the psychometrics of difference scores this would come about only when the variance of the difference score was greater than that of the pre-score (in the present case the response to control or non-critical stimuli). This condition was met in the present study as inspection of Tables 5.2, 5.14, and 5.19 indicates, but might not always be met.

The apparent conflict between results and theory that the findings with lability give rise to centres on Ben-Shakhar's (1977) dichotomization theory which was reviewed in Chapter 2. The theory was there interpreted as leading to the counter-intuitive hypothesis that stables and not labiles are more detectable. The theory postulates that subjects divide the stimuli into two classes, relevant and irrelevant, and that habituation proceeds separately for each class. The theory further postulates that frequency of stimulus presentation within a class determines habituation. Because irrelevant stimuli are more numerous than relevant stimuli in the typical test of PDD, habituation to irrelevant stimuli proceeds more rapidly and is more complete. Hence a difference in magnitude of response between the two classes of stimuli arises because of differences in habituation, and it is this difference which is responsible for detection. The theory was subsequently extended to include signal value as a further determinant of Differential Responsiveness.

Dichotomization theory was not formulated to account for findings on individual differences and its use in this way in the present research programme represented an extension of the theory. If the theory is sound, however, such an extension should be possible. It was reasoned that if habituation is a critical factor in determining Differential Responsiveness, then subjects who habituate rapidly should show larger differential responding and hence be more detectable. The individual difference dimension of lability, which subsumes a measure of speed of habituation, was thus considered the analogue of the stimulus factor of frequency explicitly referred to in dichotomization theory. The prediction clearly was not supported.

Two courses of action are suggested by the conflict: dismiss the theory, or seek to modify it. In view of the paucity of theory in the area of PDD, it would seem unwise to reject a theory that has at least some support and that is couched in theoretical concepts considered by most researchers in this area as relevant. Modification of the theory would therefore seem the better course and in fact little modification is necessary if the focus is directed more to the subsequent elaboration of the theory in terms of signal value. Although the theory is far less parsimonious with the inclusion of a further factor, it is particularly necessary in the light of the findings in Chapter 5 that the nature of the subject's verbal response to the critical question is a significant factor in detection. The study was based on the competing views of Gustafson and Orne (1965b) on the one hand and Kugelmass, Lieblich, and Bergman (1967) on the other regarding the need of the subject to actually lie to the examiner. As discussed in Chapter 5, the present results were consistent with the position of Gustafson and Orne rather than that of Kugelmass et al. The relevance of this result to dichotomization theory is that the factor of frequency of stimulus presentation was not varied from one condition to another and yet differences in responsiveness and detectability were observed. A determinant other than habituation rate is therefore implicated and it is through this other determinant that the influence of lability might be understood.

Earlier, in the discussion of the theoretical basis of lability, the more recent views of Maltzman and Bernstein on the role of stimulus significance were considered, and it was proposed that lability reflects differences in the attribution of significance to the stimulus environment. If this is accepted, and if attention is directed to the signal value component of dichotomization theory, then it is possible that labiles are more detectable because they attach greater signal value to the critical stimulus in tests of PDD. There is one difficulty with this interpretation and that is that, although labiles have been found to respond more vigorously than stables to change in stimulation following habituation (e.g., O'Gorman, 1977), they have not been found to react more vigorously to changes which are presumed to have some significance such as the subject's own name (Siddle, O'Gorman & Wood, 1979). There is thus no evidence that labiles orient more strongly to significant stimuli. The situation in tests of PDD is, however, somewhat different to that used in studies of the OR to stimulus change in that motivational level can be expected to be higher in the PDD test (e.g., competition between subject and examiner) and novelty is less in that the subject has had prior exposure to the stimului and novelty is not expressly manipulated. Under these conditions differential orienting to signal stimuli in labiles and stables may be produced.

Clearly, the results of the present study together with those of Waid, Orne, and colleagues indicate that further study of the sensitivity of stables and labiles to significant stimuli is required.

The final task of the research programme was to examine the structure of electrodermal responsiveness and the pattern of relationships between responsiveness and detectability in the context of more general dimensions of personality and temperament. Could any of the relationships observed be explained or better accounted for in terms of the major dimensions of individual differences identified by Eysenck and Eysenck (1975) and found in most self report and peer rating tests of personality? The results from this aspect of the programme were most disappointing as few significant correlations were obtained and these were not consistent with other results in the literature. For example, extraversion was not found to correlate with speed of habituation or nonspecific activity as expected on the basis of the report of Crider and Lunn (1971) nor did extraversion relate to detectability as the result of Bradley and Janisse (1981b) would lead one to expect.

Two possible explanations can be offered for this failure. First, the conditions employed here were not sufficiently similar to those employed elsewhere which led to significant relationships being observed. The use of the EPQ here rather than the EPI is a case in point. The interchangeability of these two tests as a measure of the dimensions of extraversion, neuroticism, and psychoticism has been questioned (Block, 1965). If the EPQ provides a poor measure of impulsivity (Rocklin & Revelle, 1981) and if this is the component of extraversion which mediates the relationship with lability (cf. Crider & Lunn, 1971) and possibly with detectability, then the failure to replicate previous work is not surprising. At the very least this line of argument recommends the use of both scales, EPI and EPQ, in future studies.

Second, the failure to replicate may arise from the fragile nature of the relationship. Block (1977) has pointed to the difficulties of demonstrating robust relationships between test and rating data on the one hand and physiological measures of personality on the other and even those who are cautiously optimistic about the existence of such relationships (e.g., O'Gorman, 1977) have had to admit that critics of their position can point to a substantial corpus of negative evidence.

Not too much significance can therefore be attached to the poor results with respect to personality correlates of the present project.

Finally, something needs to be said about the limitations of the present programme. The major limitation was the failure to include an innocent group in the mock agent test. Such a group would have provided information on the role of lability in predicting false positive decisions in a test of PDD. Waid and Orne (1980) employed an innocent group in their study and because no such group was included here a complete comparison of findings is not possible. An innocent group in the mock agent test would also have led to closer comparison with the results of the card test where lability as a correlate was found to be specific to subjects in the experimental group. In hindsight it would have been wiser to have employed the innocent group with the mock agent test and not include a control group the card test. This was not done at the time because of the duration of testing involved and because it was considered necessary to provide as large a sample as practicable for the multiple regression analyses which are known to produce unreliable results with small samples. Although the decision seemed sound at the time it is now recognized that a fuller answer would have been provided if an innocent group had been included.

Inclusion of an innocent group would also have helped ovecome a second limitation of the study, that of possible experimenter bias in administration of the CQ test. In the mock agent paradigm all subjects were guilty and the examiner knew this. Hence it is possible that the accuracy rates obtained in the mock agent test were reduced because of this knowledge. In hindsight it would have been more appropriate to have included an innocent group in the mock agent test and to have ensured that the subjects' guilt or innocence was unnknown to the experimenter at the time of conducting the CQ test.

In general, the results of the present research programme provide qualified support for the view that individual differences in electrodermal activity are important in laboratory studies of PDD which employ that response system. Whether this can be generalized to other response systems cannot be decided on the basis of the results reported here, although these encourage empirical attempts at such an extension. Further, the results of the present reseach programme and the conclusions from them pertain to an experiment where the motivation level of subjects was purposely kept to a minimum. Further research is required before it is permissible to generalize to deception experiments in which high levels of motivation are induced through ego involving instructions or use of monetary incentives. Whether the assessment of individual differences in responsiveness is of value to the field practioner is also impossible to judge on the basis of these data as there are sufficient differences between field and laboratory to make one cautious in extrapolating from the latter to the former.