CHAPTER 1

THE LABORATORY STUDY OF DETECTION OF DECEPTION

Attempts to detect deception by physiological means date from at least 300 BC (Trovillo, 1939). Most of the early methods, such as reaction of the skin to heat and tests of flotation, were based more on superstition or religious faith than on any physiological principles, although at least one of them, chewing rice, does seem to have a rationale in what is known of the autonomic nervous control of salivation (Orne, Thackray, & Paskewitz, 1972). With the advent in the last century of methods for measuring peripheral physiological activity in man, particularly cardiovascular functioning, techniques for the physiological detection of deception were placed on a surer footing. Lombroso is usually credited with the first application of these new methods to the detection of criminal liability (Trovillo, 1939). He employed a plethysmograph to determine changes in blood flow during interrogation. Subsequently Benussi (1914) used respiratory activity and Marston (1917) systolic blood pressure as indicators of deception, the latter proving more sensitive according to work by Burtt (1921). Larson's (1921) application of the Erlanger method of continuous measurement of cardiovascular activity and Mackenzie's development of the ink-writting polygraph in 1921 (Highleyman, 1958) marked the beginning of modern work on the psychophysiological detection of deception (PDD), work which from Keeler's first venture into the field (Barland & Raskin, 1973) has become a major commercial enterprise in the United States of America.

Application of PDD in the field has followed the approach adopted by the early reseachers. Examiners are generally trained interrogators in the military or civilian police who do not have formal training in psychology or psychophysiology. They use a portable polygraph capable of monitoring electrodermal activity, relative blood pressure, and respiration. Responsiveness in these systems during interrogation is analysed qualitatively in reaching a decision about guilt or innocence. Regardless of whether the polygraph examination involves investigation of a possible criminal act or personnel screening it is generally divided into three segments: first, a pretest interview, which may be up to one hour's duration, during which the infallibility of the "lie detector" is stressed by the examiner and specific questions are formulated with the suspect; second, the interrogation proper; and third, the post-test interview in which a subject considered suspect on the basis of the polygraph record is questioned further (Orne, Thackray, & Paskewitz, 1972).

As evidenced by recent review papers (Barland & Raskin, 1973; Grings & Dawson, 1978; Orne, Thackray, & Paskewitz, 1972; Podlesny & Raskin, 1977) PDD has become the subject of critical scrutiny by psychologists and psychophysiologists who have developed laboratory analogues of the field situation. One of the main advantages of the laboratory study of PDD is that "ground" truth, to use Podlesny and Raskin's (1977) term, can be established, i.e., the examiner knows whether a subject is innocent or guilty. In the field, ground truth cannot be precisely established despite confessions or large amounts of incriminating evidence. Other advantages of laboratory studies of PDD include the fact that laboratory equipment is more sensitive than portable field polygraphs, and the laboratory provides greater opportunities for the control and manipulation of variables. Despite these advantages a subject in the laboratory does not have the same concerns as one tested in the field. Criminal suspects realise that an unfavourable polygraph result could help to deprive them of their freedom or have other serious consequences. For the subject in a laboratory test of PDD, these consequences do not exist. Generalisations from one situation to the other must therefore be circumspect.

Two situations have been employed to study PDD in the laboratory. The first, the card test (Gustafson & Orne, 1963; 1964; 1965a; 1965b), involves the subject selecting a card from a pack and not revealing the identity of the card when questioned subsequently. Typically the subject chooses one from a set of six to eight cards and is then asked in turn about each of the cards while physiological responsiveness, usually skin resistance (SR) or skin conductance (SC) is monitored. The examiner's task is to determine which card the subject selected. A variant of this involves presenting subjects with items of personal information, such as father's name, embedded among items with no personal reference and comparing the subject's physiological reactions to the personal and neutral items. The second situation employed in laboratory tests of PDD attempts to induce greater emotional arousal and subject involvement and as a consequence simulate more closely the field situation. Subjects in this situation are requested to role play a crime, e.g., stealing money from an office draw, and to deny committing the "crime" when questioned about it subsequently (Barland & Raskin, 1975; Cutrow, Parkes, Lucas, & Thomas, 1972; Davidson, 1968; Lykken, 1959; Raskin & Hare, 1978; Podlesny & Raskin, 1978). A second group of subjects who have not role-played the mock crime are also questioned and the examiner's task is to decide on the basis of the recordings of physiological activity (e.g., SC, respiration, and cardiovascular changes) which of the subjects are guilty and which are innocent. A variant of this situation is the mock agent procedure in which the subject role-plays an espionage agent who is given a set of code words which must not be revealed during questioning. Orne et al. (1972) favour this variant of the role-play situation, as they consider it provides greater subject involvement and more experimental control.

Central to each of these approaches is the putting of a set of questions to the subject and the inferring from differences in physiological responsiveness to the questions guilt or innocence or the nature of the information that the subject has in his or her possession. Since the validity of the inferences to be drawn from differences in responsiveness depend on the nature of the questions asked, a major concern in PDD has been techniques of questioning. Four such techniques can be identified and each has been used in the laboratory.

Questioning Techniques

The most elementary form of questioning is that termed the relevant-irrelevant technique, and involves interspersing questions that are relevant to the issue about which deception is suspected with neutral or irrelevant questions. In the case of a mock crime, for example, the relevant-irrelevant technique involves asking questions not only about the crime but about other matters with no relation to the crime, e.g., "Are you sitting down?" Differential responsiveness to the relevant and irrelevant items forms the basis of an inference about guilt.

The lack of control over the arousing nature of the relevant as compared with the irrelevant questions quickly gave rise to criticisms of this technique. Reid (1947) introduced two types of control questions directed at nullifying these criticisms, viz. the comparative response question and the guilt complex question. The comparative response question is one to which the subject is expected to respond with a known or assumed lie (Abrams, 1975). Examples are: "Have you ever stolen anything?" or "Have you ever thought of stealing anything?" These questions are based on the assumption that the vast majority of people have stolen or have contemplated theft. The second type of control question introduced by Reid (1947), the "guilt complex question", is one which refers to a completely fictitious crime. Although subjects do not realise that the crime is fictitious, the examiner convinces them that they are suspects. It is assumed that subjects who are deceptive are solely concerned with relevant questions, and all other questions are inconsequential. The innocent subject, on the other hand, is thought to be more concerned about control questions to which he is deceptive. A decision about deceptiveness is based on the extent to which the relevant questions elicit greater or more complex changes than the contol questions. Because of the attempt to provide an adequate baseline against which to assess responsiveness to the critical or relevant question, this technique of questioning has been termed the control question (CQ) technique.

A decision of deceptiveness using the CQ technique is typically based (Barland & Raskin, 1975; Dawson, 1980) on a numerical index derived from evaluation of the subject's record. The respiratory, cardiovascular, and electrodermal reactions to each adjacent pair of relevant and control questions are compared. For each physiological measure a score between +3 and -3 is assigned, with the value of the score depending on the amount and direction of difference in the reactions to the pair of questions. Positive scores are assigned where the response to the control question is greater; negative scores are assigned where the response to the relevant question is greater; and a score of zero results where there is no differene in response to the two questions. Finally, scores are summed across physiological variables, pairs of questions, and trials on "charts", since it is usual to run through the question series more than once, to determine a total score. In deciding guilt or innocence this score is compared to a previously determined score range. For example, subjects with a score of +6 or greater might be classified "truthful" and subjects with a score of -6 or less might be classified "deceptive". Subjects with a score of between ±5 might be classifed "inconclusive".

The CQ test has been criticized by Lykken (1978, 1979) who argued that it is based on the belief that subjects when lying produce a specific pattern of physiological activity which is distinguishable from that produced by truthful subjects. He further argued that it is unreasonable to expect an innocent subject to show similar responses to a comparative response question and a question relating to a serious crime, particularly when responsiveness within the physiological response system, fear of the consequences of being found guilty, and confidence in the validity of the test procedure vary across subjects (Lykken, 1974). Additionally, he expressed doubt about the effectiveness of guilt complex questions. They are only adequate, Lykken maintained, if the examiner can convince the subject that he or she is suspected of both the actual and the fictictious crime. These conditions, according to Lykken, are almost impossible to fulfil in the field situation. As a consequence, he argued (Lykken, 1959, 1960), attempts to infer deception from physiological responses is an unreasonable and unwise task to set the examiner.

As an alternative Lykken proposed that physiological measures can be used to determine the significance of stimuli for subjects. This hypothesis forms the basis of the Guilty Knowledge Test (GKT; Lykken, 1959). The GKT consists of a number of multiple choice questions each containing a pertinent fact relating to the crime under investigation. Each pertinent item is included with three or four alternatives which seem equally plausible to an innocent subject who has not heard or read about the crime. The GKT is based on the assumption that a larger physiological response will be elicited by the significant alternative when presented to the guilty subject. An innocent subject, not knowing which items are significant, should not show any systematic pattern of differential responding to the items. Scoring of the GKT, Lykken claimed, is also more objective than that for the CQ test. A score of either 2, 1, or 0 is assigned to responses to the questions. The scores are determined as follows: if the physiological response is the largest to the critical item a score of 2 is assigned; if the second largest response is to the critical item a score of 1 is assigned; if the critical item fails to elicit the largest or second largest response a score of 0 is assigned. Scores are then totalled and subjects with scores above a criterion score are classified as "guilty" and subjects with scores equal to or below that score are classified as innocent.

Raskin (1978; Raskin & Podlesny, 1979) in reply to Lykken agreed that the GKT is useful and is used, but argued that the conditions under which it is appropriate greatly restrict its application. Moreover, Lykken's theoretical analysis of the CQ test, Raskin maintained, is based on two assumptions, both of which are false. They are, first, that control questions are designed to elicit truthful answers from subjects, and second, that the absence of a difference in magnitude of response to control and to relevant questions constitutes the basis of a truthful record.

On the first point, it would seem clear that Lykken, at least initially (1974), stated that the subjects are expected to answer control questions truthfully whereas the pre-test interview is designed to have the subject lie to the control questions. Lykken's position would now seem to be that the examiner cannot claim that he knows that the subject in fact lies to the control question. Raskin takes this point but maintains only that the subject is "very likely to be deceptive or concerned about them (the control questions)" (Podlesny & Raskin, 1977, p.786). On the second point Lykken maintains that he did not claim that a lack of difference in responsiveness between critical and control questions was taken as evidence of truthfulness but rather that such an outcome was designated inconclusive. The major point of contention between Lykken and Raskin would thus seem to be the extent to which control questions in fact serve as controls. Lykken maintained that innocent subjects would be more concerned about critical than control questions. Raskin countered by reporting data indicating that this is not necessarily the case (1978, p.146).

One other technique of questioning with some similarity to the GKT is the peak-of-tension (POT) procedure. Applied to the card test it involves telling the subject in advance the order in which the questions will be asked and then checking for a rise in physiological activity as a particular question approaches. In the case of the mock crime, the crucial item is typically placed third or fourth in a series of five or six alternatives which are presented in an order previously established with the subject. The criterion for the identification of deception is the pattern of physiological activity that precedes and follows presentation of the critical item. An increase in activity prior to the critical item followed by a fall as the guilty subject is presumed to relax is considered indicative of deception (Orne et al., 1972; Podlesny & Raskin, 1977).

Some idea of the value of each of these questioning techniques can be gained from the detection rates which each gives rise to, and these are reviewed in the following section.

Accuracy of PDD

Because of the differences between the card test and the mock crime, accuracy rates for these two laboratory situations are considered separately. In the laboratory, researchers have employed the card test to investigate a variety of questions about PDD, e.g., the value of multiple physiological indices (Cutrow et al., 1972), the effects of motivation (Gustafson & Orne, 1963), the method of stimulus presentation and the effects of instructional set (Gustafson & Orne, 1964; 1965a), the importance of the nature of the subjects' response (Gustafson & Orne, 1965b; Kugelmass, Lieblich, & Bergmann, 1967), and the effects of stress (Kugelmass & Lieblich, 1966; 1969). Accuracy rates (i.e., the percentage of cases in which the card selected by the subject is correctly identified from evaluation of the physiological data) for these studies vary depending on the conditions under which subjects are tested. A summary of accuracy rates for individual studies, which have employed SR or SC as the physiological measure, is presented in Table 1.1. The studies appear in chronological order. The summary focusses on electrodermal indices as these have been the widely used and the most sensitive in laboratory studies of PDD (Orne et al., 1972; Podlesny & Raskin, 1977). Inspection of Table 1.1 indicates that in the studies by Gustafson and Orne (1963, 1964, 1965a, 1965b) accuracy rates varied from 19% to 94%. The latter result occurred where subjects were presented with instructions designed to motivate deception. Following an initial card test subjects were informed that they had not been detected. On a subsequent test, a high accuracy rate was obtained. The lowest rate 19% also occurred in the second test of the same experiment

Table 1.1

Summary of Accuracy Rates for Studies Employing Card Tests

Study	Condition	Accuracy Rate	Sample Size
Gustafson & Orne (1963)	Motivated to deceive, Trial 1	67%	n=18
	Motivated to deceive, Trial 2	61%	
	Control (not motivated to deceive), Trial 1	33%	n=18
	Control (not motivated to deceive), Trial 2	22%	
Gustafson & Orne (1964)	Subject attempts to deceive examiner as to the number of the selected card (POT)	33%	n=24
	Subject attempts to deceive examiner as to whether a card selected at all (POT)	79%	
	Subject attempts to deceive examiner as to the number of the selected card (relevant-irrelevant test)	63%	n=24
	Subject attempts to deceive examiner as to whether a card selected at all (relevant-irrelevant test)	75%	
Gustafson & Orne (1965a)	Motivated to be detected	56%	n=16
	Motivated to be detected and informed of success	25%	
	Motivated to be detected	81%	n=16
	Motivated to be detected and informed of failure	88%	
	Motivated to deceive	69%	n=16
	Motivated to deceive and informed of success	19%	
	Motivated to deceive	81%	n=16
	Motivated to deceive and informed of failure	94%	

Table 1.1 continued

Study	Condition	Accuracy Rate	Sample Size
Gustafson & Orne (1965b)	Subject responds NO to all questions (POT)	74%	n=25[1]
	Subject responds NO to all questions (relevant-irrelevant test)	76%	n=27
	Subject makes no verbal response (POT)	50%	n=24
	Subject makes no verbal response (relevant-irrelevant test)	58%	n=26
	Subject generates a word association (POT)	35%	n=24
	Subject generates a word association (relevant-irrelevant test)	33%	n=23
Kugelmass & Lieblich (1966)	Minimum stress	44%	
	Medium stress	53%	n=36
	Maximum stress	47%	
	SR plus blood pressure cuff	28%[2]	n=40
	SR only	50%	
Kugelmass, Lieblich, & Bergmann (1967)	Subject responds YES to all questions	70%	n=27
	Subject responds NO to all questions	60%	
Kugelmass, Lieblich, Ben-Ishai, Opatowski, & Kaplan (1968)	SR plus blood pressure cuff	52%[2]	n=62
	SR only	56%	
Kugelmass & Lieblich (1969)	Stress condition	52%	n=62

Table 1.1 continued

Study	Condition	Accuracy Rate	Sample Size
Lieblich, Kugelmass, & Ben-Shakhar (1970)[3]	Ratio of target to non-target card is 1:2	70%	n=44
	Ratio of target to non-target card is 1:4	61%	n=44
	Ratio of target to non-target card is 1:8	52%	n=44
Ben-Shakhar (1977)	Ratio of target to non-target card is 1:8	78%	n=40
	Ratio of target to non-target card is 2:8	60%	n=40
	Ratio of target to non-target card is 4:8	45%	n=40
	Ratio of target to non-target card is 7:8	48%	n=40
	Ratio of target to non-target card is 1:2	73%	n=40
Horvath (1978)[3]	Without cuff	65%	n=20
	With cuff	46%[2]	n=20
Horvath (1979)[3]	Motivated to be detected, Trial 1	61%	n=32
	Motivated to be detected, Trial 2	56%	
	Motivated to avoid detection, Trial 1	44%	n=32
	Motivated to avoid detection, Trial 2	45%	

Notes: 1. ns in this study are not independent.

2. SR accuracy rate only.

3. Accuracy rates averaged over sex of subject and examiners.

where the subjects were again motivated to deceive but were told prior to the second test that they had been detected in the first test. If accuracy rates are aggregated across the various conditions and trials employed in the studies by Gustafson and Orne, the median accuracy rate is 62%. In the series of studies by Kugelmass and colleagues reviewed in Table 1.1 accuracy rates varied from a low of 28% to a high of 78% with a median over all conditions and trials of 52.5%, somewhat lower than the estimate based on the Gustafson and Orne series.

A procedure related to the card test involves the identification of personal information. A summary of accuracy rates for individual studies that have employed this procedure is presented in Table 1.2. The studies appear in chronological order. Across studies accuracy rates ranged from 100% to 25% with a median of 55%.

Fewer data are currently available from studies using the mock crime. Table 1.3 provides a chronological summary. Not included in the table are data from early laboratory investigations such as those of Burtt (1921), Marston (1917), Ruckmick (1938), and Summers (1939). The percentages in this table were calculated by discounting inconclusives and by using as much of the data as possible from each study. That is to say, where data were reported for "three charts only" and for "all charts" in a particular study, percentages were derived for decisions based on all charts. An exception to this is the study of Dawson (1980) for which percentages are based on only one of the four conditions employed. The condition chosen was judged to be the one most directly comparable to those used in the other studies. The other three conditions where experimental variations of the standard questioning procedure. Only slightly higher rates for the detection of innocent subjects using the CQ test would result if results for the other conditions in Dawson's study were pooled. Also included in Table 1.3 is a study by Balloun and Holmes (1979) which is the only study that did not involve a mock crime or mock agent paradigm. These researchers ensured that over half their subjects were guilty of an actual misdemeanour. The accuracy rates that are shown in Table 1.3 for their study are those for the first administration of the GKT. The second showed an overall decrease in accuracy and in the numbers of false positives.

Table 1.2

Summary of Accuracy Rates for Studies Testing for Personal Information

Study	Condition		Accuracy Rate	Sample Size
Lykken (1960)	Subjects produced voluntary autonomic responses during the polygraph test		100%	n=20
Lieblich (1969)[1,2]	Subject's name combined with five common names, unrelated to subject		71%	n=51
	Subject's name combined with five names from subject's immediate family		49%	n=53
Ben-Shakhar, Lieblich, & Kugelmass (1970)	Accuracy rates determined using mean rankings of subject's responses		78%	n=27
	Accuracy rates determined using Lykken's (1960) deviation from an assumed rectangular distribution		33%	
Lieblich (1970)[1,2,3]	Subject's name combined with five common names, unrelated to subject		25%	n=24
	Subject's name combined with five names from subject's immediate family		29%	n=24
Lieblich, Naftali, Schueli, & Kugelmass (1974)	Motivated to deceive	Trial 1	55%	n=28[1]
		Trial 10	50%	
	Motivated to deceive and consciously attempted to avoid detection	Trial 1	60%	n=20[1]
		Trial 10	42%	
	Not motivated to deceive	Trial 1	60%	n=29
		Trial 10	48%	

Table 1.2 continued

Study	Condition	Accuracy Rate	Sample Size
Ben-Shakhar, Lieblich, & Kugelmass (1975)	Ratio of personal to neutral items is 1:1	66%	n=16
	Ratio of personal to neutral items is 1:2	65%	n=16
	Ratio of personal to neutral items is 1:3	47%	n=16
	Ratio of personal to neutral items is 1:4	34%	n=16
	Ratio of personal to neutral items is 1:5	69%	n=16
	Ratio of personal to neutral items is 1:6	51%	n=16
	Ratio of personal to neutral items is 1:7	62%	n=16
Lieblich, Ben-Shakhar, & Kugelmass (1976)	Subjects were from a prison population	62%	n=30

Notes: 1. *ns* in this study are not independent.

2. Only conditions that are typically employed in PDD tests are included.

3. Subjects were children aged between 3 and 4 years.

Table 1.3

Summary of Accuracy Rates for Studies Employing a Mock Crime or Mock Agent Paradigm

Study	Measure	Test	Guilty		Innocent		Inconclusive		n
			f	%	f	%	f	%	
Lykken (1959)[1]	SR	GKT	44/50	88%	48/48	100%	-	-	49
Davidson (1968)	SR	GKT	11/12	92%	36/36	100%	-	-	48
Barland & Raskin (1975)	SR, cardio, & respiration	CQT	23/26	88%	15/21	71%	25/72	35%	72
Podlesny & Raskin (1978)[2]	SR, cardio, & respiration	CQT	15/18	83%	17/18	94%	4/40	10%	40
		GKT	7/9	78%	9/9	100%	2/20	10%	20
Raskin & Hare (1978)	SR, cardio, & respiration	CQT	21/21	100%	21/23	91%	4/48	8%	48
Waid, Orne, Cook	SR	GKT	23/29	79%	8/11	73%	-	-	40
& Orne (1978)[3]	SR	GKT	11/18	61%	9/10	90%	-	-	28
	SR	GKT	11/15	73%	12/15	80%	-	-	30
Balloun & Holmes (1979)	SR	GKT	11/18	61%	14/16	88%	-	-	34
Waid, Orne	SC	CQT	12/15	80%	12/15	80%	-	-	30
& Wilson (1979)[3,4]	SC	CQT	12/15	80%	12/15	80%	-	-	30
	SC	POT	7/15	47%	12/15	80%	-	-	30
	SC	GKT	8/15	53%	14/15	93%	-	-	30
Dawson (1980)[5]	SR, cardio, & respiration	CQT	11/11	100%	7/10	70%	3/24	13%	24

Table 1.3 continued

Study	Measure	Test	Guilty		Innocent		Inconclusive		n
			f	%	f	%	f	%	
Giesen & Rollison (1980)	SC	GKT	19/20	95%	20/20	100%	-	-	40
Waid & Orne (1980)[3]	SR	GKT	12/18	67%	9/10	90%	-	-	28
	SC	CQT & GKT[6]	9/15	60%	15/15	100%	-	-	30
Bradley & Janisse (1981a)[4]	SR	CQT	58/96	60%	56/96	58%	56/192	29%	192
	SR	GKT	57/96	59%	85/96	89%	-	-	192
Waid, Orne & Orne (1981)[7]	SC	CQT	29/40	73%	26/34	76%	-	-	74
Waid, Wilson & Orne (1981)[7]	SC	CQT	29/40	73%	26/34	76%	-	-	74

Notes: 1. Rates are based on total number of interrogations, not total n.

2. Inconclusives have been distributed equally between guilty and innocent.

3. Lykken's ranking method not employed in evaluating results.

4. ns in this study are not independent.

5. Only data for condition IAT included as this more directly comparable with other studies. Estimates based on DAT question would have yielded higher accuracy for innocent subjects.

6. Detectability assessed across two CQTs and one GKT.

7. Detectability assessed by comparing mean SCR to relevant questions across three tests with mean SCR to control questions across three tests.

Excluding Waid and Orne's (1980) method of assessing detectability across two CQ tests and one GKT (see Table 1.3), the median accuracy rate for the CQ test from Table 1.3 is 80% for guilty subjects and 76% for innocent subjects, and in most studies there is a substantial (10%+) inconclusive rate. The median accuracy rate for the GKT is 73% for guilty and 90% for innocent subjects. If attention is confined only to those studies which employed Lykken's method of scoring, the median rates are 83% for guilty and 100% for innocent. Only one study using the GKT reported a substantial number of inconclusives.

In general, detection rates are high in studies in which the mock crime is used. This agrees with Thackary and Orne's (1968) estimates that the accuracy rates for the mock crime were of the order of 50% to 94%, whereas those for more neutral material (e.g., cards) were of the order of 40% to 83%. The difference was attributed to variations in motivational level of the subjects. This raises the question of the basis of such effectiveness as PDD has been shown to have, a question taken up in the following section.

Theories of PDD

The basis for the effectiveness of the laboratory card test is unclear, as theories of PDD developed by field workers do not provide a satisfactory explanation. Davis (1961), for example, in his review of theories of PDD indicated three major types of explanation: conditioned response, conflict, and threat of punishment. According to the conditioned response theory the relevant questions act as conditioned stimuli associated with the subject's involvement in a criminal activity, and as such evoke emotional responses in the subject. The conflict theory states that subjects are socialized to tell the truth and that physiological disturbances result when the subject behaves in a way contrary to the press of socialization. The threat of punishment theory argues that deceptive subjects produce large physiological responses because they fear the consequences of being detected. None of these theories is consistent with findings of the laboratory application of the card test. For example, a laboratory card test does not involve the level of emotion implied by the conditioned response theory or the consequences of the punishment theory. The conflict theory, on the other hand, is not consistent with the findings of Kugelmass, Lieblich, and Bergmann (1967) that accuracy was significantly greater than chance even though subjects were not required to lie. The validity of these theories as applied to PDD in the laboratory is thus in doubt.

Attempts to provide a more satisfactory theoretical understanding of the card test have been made by Gustafson and Orne (1963, 1965a) and by Ben-Shakhar (1977). Gustafson and Orne assumed with previous theorists that a motivational factor was important but proposed that in the context of the laboratory card test it is not fear of the consequences that motivates behaviour but desire to succeed. What is and what is not a success depends on how the experimenter structures the task so that being detected may be as important under some circumstances as not being detected is under others. The results of Gustafson and Orne (1963, 1965a) are consistent with this view.

The approach of Ben-Shakhar is more radical. He proposed, at least initially (Ben-Shakhar, Lieblich, & Kugelmass, 1975), that the effectiveness of the card test was due to two processes. The first is the categorization by the subject of the stimuli (card numbers) into two classes, one containing the selected or relevant stimulus and the second containing all other stimuli. The second process is that of habituation of autonomic responses to stimuli within each class, which takes place on stimulus presentation. Habituation, he maintained, occurs independently for each class and generalizes to all stimuli within a class. The variable influencing habituation according to Ben-Shakhar is the frequency of stimulus presentation. Since relevant stimuli are typically presented far less frequently than irrelevant stimuli in the card test, habituation to relevant stimuli is less and hence response magnitude to them is greater.

The theory rests in part on findings reported in the literature on the orienting reflex (OR). The OR is a non-specific reflex which is initiated by a qualitative change in a stimulus and is subject to the process of habituation (Sokolov, 1960). In addition to behavioural responses, the OR includes cortical and autonomic components. If a stimulus is repeatedly presented, the OR that it elicits will decrease in magnitude (Mackworth, 1969; Lynn, 1966) i.e., the OR "habituates". According to Sokolov (1963) this is the result of formulation of a neural model of the features of the stimulus in the cortex. If a new, different, or novel stimulus is presented the OR reappears as the characteristics of the new stimulus fail to match the neural model of the original stimulus. Applied to the card test, OR theory interprets the autonomic responses associated with presentation of stimuli (card numbers or questions) as ORs.

Evidence for Ben-Shakhar's dichotomization theory, as he termed it, is drawn from studies conducted by Ben-Shakhar and colleagues in which the frequency of occurrence of stimuli within the relevant and neutral class is manipulated and its effects on the efficiency of detection examined (Ben-Shakhar, 1977; Ben-Shakhar, Lieblich, & Kugelmass, 1975; Lieblich, Kugelmass, & Ben-Shakhar, 1970). The results of Ben-Shakhar (1977) are particularly interesting in this respect. In a condition in which the ratio of relevant to neutral cards was 7:1 the accuracy rate was a low 48%. As predicted, "negative detection" was observed, i.e., subjects were more responsive to the neutral stimulus than to the relevant stimuli. Second, conditions which varied in the frequency of presentation of neutral stimuli did not lead to differences in the accuracy of detection. A group presented with a series containing seven neutral cards and one relevant card were no more poorly detected than a group presented with a series containing only one neutral and one relevant card. According to Ben-Shakhar habituation generalizes to all stimuli within the class.

Ben-Shakhar's theory, it should be noted, attempts to account for differential responding in the card test (as well as a variety of other paradigms not relevant to the present discussion) without recourse to a motivational factor. The explanation is simply in terms of information processing in which stimuli are categorized and separate neural models are constructed for separate classes of stimuli. At least this was the earlier and stronger version of the theory. Ben-Shakhar considered the possible influence of motivational factors in terms of the role of signal value of a stimulus on OR magnitude and habituation to that stimulus, but initially discounted it on the grounds of parsimony and failure to observe differences in habituation to signal and non-signal stimuli (Ben-Shakhar, Lieblich, & Kugelmass, 1975). According to Sokolov, stimuli with signal value because of their biological significance (the squeal of a mouse to a cat, for example), history of association with other stimuli (e.g., the conditional stimulus in classical conditioning), or because salience is conferred on them by experimental demands (e.g., instructions to pay attention) elicit more pronounced ORs than neutral or non-signal stimuli. In these terms the selected card in a card test has signal value because it, rather than any other, was selected by the subject. Increased responsiveness to the chosen card could thus be interpreted as a consequence of its signal value. Initially, Ben-Shakhar rejected this interpretation as unnecessary but as a result of subsequent work recognized it as a possibility, since the differentiation between conditions varying in frequency of stimulus presentation was not as clear as the original model would predict (Ben-Shakhar, 1977). Ben-Shakhar's position would now seem to include signal value as a factor. Differential responsiveness in the card test, according to the theory, is a function of the subject's categorization of stimuli into two classes, relevant and neutral, and the differences in orienting activity that these two classes occasion. Orienting to stimuli in the relevant class is more sustained because stimuli in this class are fewer and also because they have greater signal value. Such a hypothesis is consistent with several current views of the OR (e.g., Bernstein, 1979; Maltzman, 1979).

In the case of the mock crime, theories of PDD developed in the field would be expected to have greater application than is the case with the card test. The tendency, however, has been to interpret the effectiveness of PDD in the mock crime within much the same framework of OR theory as that used by Ben-Shakhar.

Lykken (1974) proposed that the effectiveness of the GKT was due to the elicitation of an OR to the significant stimulus. He argued that all subjects tested on a GKT would be expected to produce an OR to each alternative in the test. Guilty subjects, who recognise the pertinent item, can be expected, however, to produce a stronger OR to it because of the increased signal value for them.

More recently, Raskin (1979) applied the concepts of orienting and defensive responses to an analysis of the effectiveness of both the CQ test and the GKT. He argued that the importance of the pretest interview in the CQ test is to focus the subject's attention on the control question and away from relevant questions. As a consequense the control questions have more signal value than the relevant questions for the innocent subject. Not only do the innocent subjects find the relevant questions less significant because they are innocent, but the non-specific nature of the control question, and the thought that subjects must give to them if they are to be answered "correctly" lead to information processing which enhances and maintains the responses of innocent subjects to them. According to Raskin, signal value and information processing make the responses of the innocent subject to control questions less resistant to habituation. To the guilty subject, on the other hand, the control question is perceived as less significant than the relevant question. Hence it produces a smaller response and more rapid habituation. The relevant question has a great degree of signal value for the guilty subject. Hence it produces stronger physiological responses and more resistance to habituation.

In the GKT, according to Raskin, there is minimal information processing for all subjects. The pertinent alternative has no signal value for innocent subjects . Hence the signal value is equally low for all alternatives. With the use of multiple questions the chance of an innocent subject producing an OR to a series of pertinent alternatives is extremely small. For the guilty subject all pertinent alternatives have signal value and they produce ORs greater than the response to non-pertinent alternatives.

Raskin (1979) considered that the dichotomization theory of Ben-Shakhar could be applied to interpretation of the GKT. Two classes of stimuli are identified by the guilty subject and habituation to the relevant items is less than to the irrelevant items. Raskin, however, argued for explicit inclusion of a motivational factor in the theory on the grounds that the higher detection rates with mock crimes than with card tests indicate the importance of subject involvement and motivation.

In the case of the CQ test Raskin extended the interpretation in terms of OR theory to include reference to defensive reflexes (DRs) as well. DRs, according to Sokolov (1963), are produced by high intensity stimuli. Like ORs, DRs may involve behavioural, cortical, and autonomic responses, but unlike ORs they habituate slowly if at all. As a further basis for differentiation Sokolov proposed that the OR was characterized by vasodilation in the forehead whereas the DR was characterized by vasoconstriction. This patterning has, however, proved difficult to demonstrate (e.g., Skolnick, Walrath, & Stern, 1979).

Raskin (1979) used data on cardiovascular changes to support his interpretation in terms of ORs and DRs. Analysis of heart rate (HR), diastolic blood pressure (DBP), and finger pulse amplitude (FPA) over a 20 sec. period following the presentation of relevant questions indicated that guilty suspects reacted with a rapid HR acceleration and vasoconstriction which produced a rapid rise in the level of DBP. The rise in DBP activated a baroreceptor reflex which produced rapid HR deceleration, which was associated in turn with a fall in the level of DBP. The latter was not, however, as substantial as the fall in HR, possibly because peripheral resistance was still increasing and this limited the DBP change. Following the HR deceleration, acceleration occurred for a second time and DBP rose accordingly. Raskin argued that this pattern of cardiovascular responses is consistent with the view that the relevant questions elicit a DR rather than an OR. Acceptance of this hypothesis must be deferred, however, until the nature of the DR is better understood and Raskin's observations are replicated.

The Question of Individual Differences

As this brief review of the literature on PDD in the laboratory has indicated, decisions about deceptiveness, which in general have a reasonable degree of accuracy, depend on the subject showing differential physiological responsiveness to relevant and control questions selected and presented according to one of a number of now standard techniques. Differential responsiveness is thought to be due to differences in orienting to the two classes of questions which in turn is accounted for in terms of habituation and the effects of signal value. This view has evolved from consideration of the influence of a variety of stimulus characteristics and situational factors on differential responsiveness in studies using card tests or, more recently, mock crime or mock agent procedures.

In contrast to the amount of attention devoted to stimulus and situational factors on differential responsiveness, individual difference variables have been relatively neglected. Barland and Raskin (1973) in their review of detection of deception pointed to detectability in various populations as one of the major problems requiring research: "...it is apparent" they argued "that the psychological composition of a given S at any particular moment is critically important in determining how and to what extent that S will respond" (Barland & Raskin, 1973, p.454), and went on to enumerate a number of possible individual difference factors, such as degree of tendency to lie, psychopathy, neuroticism, psychoticism, age, and IQ, and the like that might be investigated. Lykken (1974) argued that individual differences in fear of the consequences of detection, autonomic lability, and confidence in the effectiveness of the procedures were important sources of confounding in traditional approaches to PDD, and singled out fear of the consequences as most significant in questioning procedures other than the GKT. More recently, Podlesny and Raskin (1977) in their review noted that one of the major methodological problems in PDD turned on the nature of individual differences: "The facts that some individuals are highly responsive but others are unresponsive, that some show maximum responsiveness in one physiological system but others show minimum responsiveness in that system, and that some individuals manifest a great deal of change in ongoing activity between stimuli but others show little or no change between stimuli challenge those who would attempt to reach conclusions concerning deception in the specific case" (Podlesny & Raskin, 1977, p.785).

In this passage Podlesny and Raskin draw attention to three aspects of physiological responsiveness of possible importance to PDD: the absolute responsiveness of the system; its susceptibility to change; and the relative responsiveness of the system in relation to others. These distinctions have a substantial basis in the psychophysiological literature; responsiveness and susceptibility to change have been discussed, particularly with reference to the electrodermal system, in terms of factors of reactivity (e.g., Martin & Rust, 1976) and lability (e.g., Crider & Lunn, 1971), and the idea of relative reactivity is summed up in the concept introduced by Lacey and Lacey (1958b) of individual response specificity (IRS). Although some attempt has been made of late to examine some of these features of responsiveness in relation to PDD (Waid & Orne, 1980), there has been, as Waid and Orne acknowledge, little systematic investigation of the problem. Such an investigation would require: (a) identification of the basic dimensions of individual differences in responsiveness, (b) consideration of their relationship to less proximal factors such as psychopathy and anxiety which have been implicated on intuitive grounds in the problem of detectability, (c) explication of the ways in which these dimensions might be expected to relate to differential responsiveness, and, finally (d) the testing of these relationships empirically. The research programme described in the chapters that follow was directed to investigation of each of these points with respect to the electrodermal system, since this system has been most widely used in laboratory studies of PDD and is one about which a good deal is known.

CHAPTER 2

INDIVIDUAL DIFFERENCES AND DETECTABILITY

The purpose of the present chapter is to review research bearing on the basic dimensions of individual differences in electrodermal activity and the relationship of these dimensions to more general dimensions of individual differences as a prerequisite to consideration of individual differences in PDD. The review is organized around three topics, the first two of which are concerned with studies of individual differences in general, and the third with individual differences in relation to PDD.

Electrodermal Reactivity and Lability

A recording of electrodermal activity, whether it is in terms of SC, SR, or skin potential (SP), can be analysed in a number of ways (cf. Martin & Venables, 1980), but the most widely employed are in terms of level, magnitude of response to stimulation, and frequency of nonspecific responses (NSRs), i.e., responses similar in wave-form to those evoked by stimulation but not time-locked to any identifiable stimulus (Lacey & Lacey, 1958a). To these could be added the various time measures such as latency and recovery time, and the measures of change in the response system as a consequence of stimulation as, for example, in classical conditioning. Among the latter group of measures are those concerned with habituation (response decline with repeated stimulus presentation), which have been extensively investigated in recent years because habituation is a defining feature of the OR. O'Gorman (1977) identified five types of measures of habituation, with the most frequently employed being number of stimulus presentations to reach a criterion of habituation such as two or three consecutive response failures, the frequency of responding during a series of stimulus presentations, and the change in magnitude of response over trials as estimated by the slope of a regression line. The first two of these were the most widely used, according to O'Gorman's (1977) review, and were found to correlate highly.

The meaning of these various measures of electrodermal activity, their interrelationship, and their correlation with personality and psychopathology have been the subject of considerable study (see Prokasy & Raskin, 1973 for a detailed review to that date). The most frequently employed concepts for predicting and explaining these relationships have been arousal, orienting and attention, anxiety, and extraversion. The brief review which follows attempts to establish what dimensions of individual differences have been identified as a consequence of research to date and what concepts can be best used to describe them.

The level of electrodermal activity, which in the case of SC typically falls as the subject relaxes and rises with task demands, was considered by Duffy (1962), Malmo (1962) and other arousal theorists as an index of the level of central arousal or activation. Raskin (1973) in his review of the question concluded that this view of electrodermal level was well supported by the earlier evidence summarized by Woodworth and Schlosberg (1954) as well as by more recent data. The conclusion must be qualified, however, by the findings that individual differences in electrodermal level do not correlate with other measures presumed to index central arousal such as HR level and muscle activity (e.g., Mathews & Lader, 1971). It would thus be more correct to describe electrodermal level as an index of activation of the electrodermal system.

Magnitude of the electrodermal response to a stimulus might be expected to index electrodermal activation since according to Wilder's law of initial values (LIV, Wilder, 1957) magnitude of response is related to the level from which it arises. As formulated by Wilder, the LIV states "Not only the intensity but also the direction of a response of a body function to any agent depend to a large degree on the initial level of that function at the start of the experiment. The higher this 'initial level', the smaller is the response to function-raising, the greater is the response to function-depressing agents. At more extreme initial levels there is a progressive tendency to 'no response' and to 'paradoxic reactions', i.e., a reversal of the usual direction of the response" (Wilder, 1957, p.73). When applied to electrodermal activity, the LIV would predict a negative correlation between level and response magnitude, i.e., high levels should be associated with small response magnitudes. Some support for this was reported by Lacey (1956) from a series of studies conducted in his laboratory when magnitude of response to the stimulus was expressed as a percentage of the prestimulus level. For magnitude of response expressed as the difference between pre- and post-stimulus level, the trend of his results was opposite to that expected from the LIV; for 10 of the 12 samples response magnitude was positively correlated with prestimulus level and significantly so in five of the studies (rs of from .31 to .57). Significant positive correlations (rs of .35 to .77) between prestimulus level and magnitude of the SCR (skin conductance response) were also reported for three samples by Hord, Johnson, and Lubin (1964).

Thus there is little evidence for the applicability of the LIV to electrodermal data expressed in conductance units, a conclusion supported by a careful intra-subject study of the question conducted by Bull and Gale (1974). The positive correlation between level and response magnitude suggests, however, that the converse of the LIV may hold, and this in fact was the conclusion of Hord et al. (1964). Further support for this is to be found in the data from a number of studies summarized by Venables and Christie (1980) which demonstrate correlations of the order of .24 to .62 for samples with a combined n of over 2000.

The positive correlation between magnitude and level suggests that both might be considered indices of electrodermal activation, with level reflecting the extent to which the individual will respond if stimulated, i.e., the subject's reactivity within the electrodermal system. There is, however, an alternative, and possibly conflicting interpretation of the magnitude of response offered by Maltzman and Raskin (1965). They employed individual differences in magnitude of SC to a simple sensory stimulus as an index of the extent to which individuals orient to a stimulus, and observed a pattern of relationships with this measure which they argued was better described in terms of attentional than arousal processes. Maltzman and Raskin specifically rejected an interpretation of response magnitude in terms of central drive state, a concept similar to that of central arousal level.

Such a conflict is also discernible in the competing interpretations of frequency of NSRs. The sensitivity of this measure to drugs presumed to influence central arousal level (Burch & Greiner, 1958) and to variations in stress on the subject (e.g., Cohen, 1967) point to an arousal interpretation of this measure. Such an interpretation has been offered by Lader and Wing (1966) who found NSR frequency a sensitive measure of arousal in clinical samples. The failure of NSR frequency to correlate consistently with electrodermal level was noted by Crider and Lunn (1971) who maintained that NSRs were more likely related to shifts in attention than to general arousal. Studies linking NSR frequency with performance in vigilance tasks (see Katkin, 1966) provide support for an attentional interpretation of this measure.

A further line of evidence supporting an attentional view of NSR frequency is the close correlation of this measure with measures of habituation, particularly those based on frequency of evoked responses in a stimulus series or number of stimuli to an habituation criterion (Bohlin, 1973; Coles, Gale, & Kline, 1971). The replicability of the relationship between speed of habituation and NSR frequency led Crider and Lunn (1971) to propose that both measures indexed a dimension of individual differences which they termed electrodermal lability. The term lability was used by Lacey and Lacey (1958a) to characterize individual differences in NSR frequency. They described subjects who show low frequencies of NSRs as "stables" and subjects showing high frequencies as "labiles". The stable/labile classification was found to have reasonable stability over time and consistency across experimental situations, an observation replicated by O'Gorman and Horneman (1979). The addition of speed of habituation as a further basis for definition of lability was justified by Crider and Lunn (1971) on the grounds of the equivalence of the two measures in terms of stability over time, their high intercorrelation, and their relationship with a self report measure of the personality dimension of extraversion.

One method of examining the interrelationship of a number of measures is factor analysis. The aim of factor analysis is to summarize the pattern of correlations among a set of variables in a way which is economical and which allows the original matrix of correlations to be reproduced (Nunnally, 1967). Factor analysis was used by Martin and Rust (1976) to study the structure of the electrodermal system. They employed two samples of male subjects. The first consisted of 149 prisoners and 63 control subjects, and the second consisted of 42 pairs of twins. All subjects were presented with an habituation task which consisted of 21 auditory stimuli, of 95 dB, and 1000 Hz. Thirteen measures common to both samples were derived for analysis. These included SC level, magnitude, latency, recovery time, and the changes in these variables during the stimulus series estimated from regression equations, and counts of the number of stimulus evoked responses and NSRs. The 13 measures plus two or three unique to each sample were factor analysed separately for each sample and the resulting factors subjected to oblique (Promax) rotation. For the larger of the two samples, five factors were found to account for 78% of the variance. The first of these was defined by mean level and mean response amplitude. The second factor, which correlated less than .1 with the first was defined by change in level and change in response amplitude. The third and fifth factors were also defined by measures of change in latency and in recovery time. The fourth factor was defined by mean values for latency and recovery time, and was correlated -.47 with the first factor. The number of evoked responses correlated with both the first and the second factors, but with the second more strongly, and the number of NSRs correlated with the second and fourth factors. This factor structure was not replicated in all respects in the second sample, although a response magnitude and level factor, a change in magnitude and change in level factor, and a factor loading the time-based measures emerged.

Martin and Rust (1976) interpreted their findings as support for "general reactivity as a common element in SCR variables" (p.559). O'Gorman (in press), however, pointed to the separation revealed in their analyses between magnitude and level measures on the one hand and measures of change in response or habituation on the other, and argued that this was consistent with the interpretation offered by Crider and Lunn in that the factor of response change loaded the measures of number of evoked responses and frequency of NSRs. That is, a reactivity factor (Factor 1 in sample 1 of Martin and Rust's study) could be distinguished from a lability factor (Factor 2 in their study). This interpretation receives some support from the factor analyses of SC measures reported by Lockhart and Lieberman (1979). In contrast to the measures employed by Martin and Rust which were based on responses over 21 trials, those derived by Lockhart and Lieberman were based on response to one stimulus only. As such, measures of change in response (habituation) were not included in the analyses, which were conducted on samples of hospital staff (_n_ = 40), schizophrenic patients (_n_ = 20), and a group of mentally retarded individuals (_n_ = 20). Three factors were extracted in each sample and rotated to an orthogonal solution (Varimax). For the non-patient sample, the first factor was defined by response amplitude and loaded level. The second factor, uncorrelated with the first because of the nature of the rotation, was defined by latency and loaded "elasped time" which, as their description indicates, is an indirect measure of NSR frequency. Thus the factor which best described the index of lability available in the study was independent to that common to response amplitude and level.

Although less than clear-cut the results of the factor analyses and the pattern of intercorrelations arising in other studies point to at least two factors being responsible for variance in measures of electrodermal activity. The first, termed reactivity, is defined by level and magnitude measures and can be thought of as a function of arousal or activation. The second, termed lability, is defined by measures of habituation and NSR frequency and can be thought of as a function of the maintenance of orienting or attention. Such a two-factor solution is consistent with a range of two-process models currently discussed in the psychophysiological literature, e.g., the stimulus comparator and arousal model of Sokolov (1969), the arousal modulation and tonic arousal model of Claridge (1967), and the sensitization and habituation model of Thompson (Thompson, Groves, Teyler, & Roemer, 1973). There are of course important differences among these models but each postulates that the final response output to stimulation is partly a consequence of the state of the system (arousal, tonic arousal, sensitization) and partly a consequence of stimulus processing (stimulus comparison, modulation, or habituation). The two factors of reactivity and lability may be considered the structural counterparts of these two processes, with reactivity reflecting individual differences in the state of the system and lability reflecting individual differences in stimulus processing. The more reactive subject might be thought of as more aroused and the labile as engaging in more processing of the stimulus. Although speculative, these considerations indicate that a two-factor solution is not an unexpected result.

The personality correlates of these various dimensions of electrodermal activity has been the subject of investigation by a large number of researchers. As the reviews of O'Gorman (1977; in press) and of Orlebeke and Feij (1979) make clear, however, there are few replicable findings. Among the earliest studies are those of Mangan and O'Gorman (1969), Crider and Lunn (1971), and Coles, Gale, and Kline (1971).

Mangan and O'Gorman (1969) conducted two experiments on SRR (skin resistance response) to a series of tones of moderate intensity. In the first experiment 20 male subjects were divided into one of the following four categories on the basis of their score on the neuroticism (N) and extraversion (E) scales of the Eysenck Personality Inventory (EPI, Eysenck & Eysenck, 1964): high N-high E, high N-low E, low N-high E, and low N-low E. In the second experiment 24 male subjects all with low scores on N were divided into low E (introvert) and high E (extravert) groups, again on the basis of the EPI. These researchers were interested in the relation between personality and the amplitude of the initial response and the number of trials to an habituation criterion. They reported a complex interrelation between personality and the SRR variables in the first experiment. For initial amplitude, subjects with high N scores produced smaller responses than low N subjects. This held regardless of the level of E. For speed of habituation high N-high E subjects and low N-low E subjects habituated more slowly than high N-low E and low N-high E subjects. In the second series extraverts produced a greater initial response than introverts and were more rapid habituators.

Crider and Lunn (1971) recorded SPR (skin potential response) to a series of tones for a sample of 22 male subjects. The researchers were specifically interested in frequency of NSRs and speed of habituation to a criterion of three consecutive response failures. Subjects completed the Minnesota Multiphasic Personality Inventory (MMPI, Hathaway & McKinley, 1967) to provide data on the following personality variables: neuroticism, extraversion, and impulsivity. Extraversion was found to be significantly correlated with the rate of NSRs ($r = -.38$, $p < .05$) and with speed of habituation ($r = -.48$, $p < .05$). Neuroticism, however, failed to correlate with either measure. Four scales of impulsivity were employed and speed of habituation was found to correlate significantly with all of them. Only two of the four scales significantly correlated with frequency of NSRs.

Coles, Gale, and Kline (1971) studied SCR to 20 tone stimuli for male subjects who were divided into three levels of extraversion (high, medium, and low) and two levels of neuroticism (high and low) on the basis of their scores on the EPI. A number of measures of reactivity were derived. They included SCR magnitude to the first stimulus, SC level, NSRs, and three habituation measures. These researchers reported that subjects with high scores on neuroticism were more labile in terms of the three measures of habituation. However, total number of NSRs was inversely related to extraversion. There was no relationship between personality group and either SC base level or SCR magnitude.

More recent studies of the personality correlates of electrodermal reactivity and lability suggest that stimulus factors may moderate relationships. Fowles, Roberts and Nagel (1977) conducted a series of studies into the possible relationship between SC level and extraversion as measured by Block's (Block, 1965) Ego Control scale. In their first two studies they divided a group of undergraduate male subjects into high E and low E subjects while keeping neuroticism relatively constant. Subjects performed either an easy or a difficult task prior to being presented with either a series of 20 high (103 dB) or low (83 dB) intensity tones. In their third study, Fowles et al. divided a group of female subjects into high, medium, and low E categories. A rest period rather than a task preceded the presentation of the tone series. Although a 83 dB tone series and a 103 dB tone series were employed subjects were not exposed to both series. In their fourth study a group of female subjects were divided into one of the four possible combinations of high E, low E, high N, and low N. One of two sets of stimulus intensities (75 dB or 100 dB) were presented to subjects.

The results of the study indicated that for subjects in the difficult task condition, extraverts demonstrated a large increase in level in response to an increase in stimulus intensity whereas introverts showed no increase. Further, the level for extraverts who were presented with high intensity tones failed to decline over trials. The results of the studies that included a rest period indicated that introverts demonstrated higher levels than extraverts. This finding was most marked for the lower tone intensities, i.e., 75 dB and 83 dB tones as opposed to 100 dB and 103 dB tones. On the basis these findings Fowles et al. (1977) suggested the empirical generalization that introverts demonstrate high SC levels at low to moderate stimulus intensities, and that extraverts demonstrate high SC levels at high stimulus intensity.

Smith and Wiggleworth (1978) divided 162 subjects into one of nine groups depending on their scores on both the E and N scales of the EPI. High, medium, and low levels E were thus combined with high, medium, and low levels of N. SR was recorded while subjects were presented with a series of 60, 80, and 100 dB tones of 1000 Hz. The presentation of these blocks of tones was counterbalanced across subjects in each personality group. Within each series of tones the criterion for habituation was "two successive responses of less than 1000 ohms each" (p.287). Once this criterion had been met a 3300 Hz test tone of the same intensity was presented. To test for dishabituation one extra standard stimulus followed the test stimulus. For speed of habituation, Smith and Wigglesworth (1978) reported that subjects with N scores in the medium range were significantly slower habituators than subjects with either high or low N scores. Further, they found that at low (60 dB) and medium (80 dB) intensities there were no differences in habituation due to extraversion but that at the high intensity (100 dB) subjects with high E scores were rapid habituators. Analysis of response magnitude was more equivocal with effects for E and N being reported depending on the particular index used.

Taken together, these various studies indicate little evidence of a consistent relationship between measures of either the reactivity or lability indices and the personality dimensions of extraversion and neuroticism. Although the vast majority of studies examining correlates of electrodermal activity have employed these dimensions, one other which deserves mention because it has used in a number of studies of PDD is that of socialization as measured by the California Personality Inventory (CPI, Gough, 1964). The relation between socialization and electrodermal activity is documented in two studies. Waid (1976) divided a sample of 48 male and 32 female students into groups with high, medium, and low scores on CPI socialization. SCR was recorded while subjects were presented with 30 pairings of a warning signal followed by a 98 dB noise and 30 pairings of a signal followed by a 68 dB tone. Waid reported significantly smaller SCRs to noise for the low socialization group than from the high group. There was no significant difference, however, between SCR amplitude to the innocuous tone for the two groups. In a later study Waid, Orne, and Wilson (1979) recorded SCR for 26 male subjects who were presented with an unexpected startle stimulus, a loud hand clap. These subjects also completed the socialization scale of the CPI. Consistent with the findings of Waid (1976), subjects with low scores on socialization produced significantly smaller SCRs to the loud startle stimulus than subjects with high scores. The correlation between SCR and socialization was .46 ($p < .01$). Thus socialization would seem to relate to at least one index of reactivity under conditions of intense stimulation.

Individual Response Specificity

As noted in Chapter 1, individual differences in physiological responsiveness can arise not only within response systems but also between response systems. That is, subjects may differ in terms of which of a number of response systems are their most reactive. Some subjects, for example, may be "skin conductance responders" in that they respond maximally in terms of SC rather than in terms of HR or respiratory change. This observation was first made by Lacey, Bateman, and Van Lehn (1953) who examined Malmo's "principle of symptom specificity" (Malmo & Shagass, 1949; Malmo, Shagass, & Davis, 1950) in a non-clinical sample. Malmo had found that patients with symptoms in a particular organ system, e.g., headache, cardio-vascular disturbance, were more physiologically reactive in that particular organ system. Lacey et al. (1953) reformulated the principle of symptom specificity as follows: "for a given set of autonomic functions, individuals tend to respond with a pattern of autonomic activation in which maximal activation will be shown by the same physiological function, whatever the stress" (Lacey et al., 1953, p.8). To test this principle, they exposed 85 male college students to four tasks (mental arithmetic, hyperventilation, letter association, and cold pressor) and monitored base level and stress levels of SC, HR, and variability of heart rate (VHR). The extent to which the subject reacted to stress in SC, HR, and VHR was measured in two ways. First, what the researchers' termed an Autonomic Tension Score (ATS) was derived for each response system. This was a measure of the level the physiological variable reached during stress. The second was what they termed an Autonomic Lability Score (ALS) for each system and represented the "displacement" the physiological variable showed under stress, when adjusted for any dependence of stress level on prestimulus level (the LIV). Both the ATS and ALS scores were essentially standard scores that expressed responsiveness in terms of sample variability and hence permitted comparison across response systems.

In the analysis of these data Lacey et al. (1953) found varying support for their version of the specificity principle. Some subjects showed "maximum" specificity in that one physiological variable was most responsive in all conditions. For other subjects specificity was only what they described as "high" in that the one physiological variable was most responsive in three of the four conditions. For still other subjects specificity was either "low" (where one physiological variable was most responsive in two conditions and another physiological variable was most responsive in the other two conditions) or "minimal" (where one physiological variable was most responsive in two conditions and the other physiological variables were most responsive under different and separate conditions). Overall, however, there was sufficient support for the generalization that for a number of subjects the distribution of reactivity across systems is non-random. As well as examining simply the most reactive systems, Lacey et al. (1953) studied the pattern of activation across systems. With three physiological variables six patterns of response are possible (PC > HR > VHR; PC > VHR > HR; HR > PC > VHR; HR > VHR > PC ; VHR > PC >HR; and VHR > HR > PC). Given four stresses a particular hierarchy may be exhibited under only one or up to all four. Analysing their data for the maintenance of a hierarchy of responding, Lacey et al. (1953) found that some individuals reproduced the pattern of response irrespective of the nature of the stress, and this was true whether the ATS or ALS measure was employed. As this was a stronger demonstration of the original principle, Lacey et al. (1953) restated it in the following way: "For a given set of autonomic functions quantitative variations among individuals exist in the degree to which a pattern of response is stereotyped" (p.20). That is, some subjects will display response stereotypy whatever the stress. Others will show greater variations across stress conditions, but one pattern of response will dominate. Other subjects will randomly exhibit different patterns of specificity across stress conditions.

In a subsequent study Lacey and Lacey (1958b) varied their subject sample, dependent variables, and conditions. Forty-two adult women were exposed to four conditions (relaxation, cold pressor, mathematics, and word fluency) while systolic blood pressure (SBP), diastolic blood pressure (DBP), SC, HR, VHR, and pulse pressure (PP) were recorded. Both ATS and ALS scores were computed for stress and anticipation periods. Specificity was examined in terms of the concordance of the rank order of scores for systems across conditions, using Kendall's coefficient of concordance. For ATS, the coefficients of concordance for 93% of subjects were significant at the .05 level, and 73.8% were significant at the .01 level. For ALS, 21.4% of the coefficients of concordance were significant at the .05 level.

Lacey and Lacey (1962) also reported on the consistency of individually exhibited response patterns over a four year period. Complete data were available on all physiological variables in both test and retest sessions for 20 boys and 17 girls. Physiological variables that were recorded were: SC, HR, VHR, SBP, and DBP. Only one stimulus condition, cold pressor, was employed but various phases of the task including anticipation and recovery were examined. The several analyses of these data reported by the Laceys supported the hypothesis that stereotypy of response is an enduring characteristic of the subject.

The essential findings of this programme of studies by the Laceys and their colleagues can be summarized as follows: 1. Individuals show maximum responsiveness in one particular response system (individual response specificity) and that the entire pattern of responsiveness in a group of response systems is reproducible over situations (individual response stereotypy). 2. That there are quantitative variations among individuals in the extent to which stereotypy (as summarized by the coefficient of concordance) is expressed and that these variations show considerable stability over time, i.e., stereotypy can be considered an individual difference variable. 3. The phenomena are demonstrable whether response levels or change scores (corrected for dependence on initial levels) are employed, although there is some suggestion that the effect is more likely to be obtained with response level measures.

What this work does not indicate is the frequency of subjects who can be expected to react most strongly in any particular response system. This of course would require a norming study of some magnitude and this question was not the concern of the Laceys. More importantly, the concepts of specificity and stereotypy refer to relative effects in that the system which is found in any particular study to be the most responsive is only most responsive in comparison to the others included in the study. Although the number of possible response systems that could be included is finite, it is usually only possible to study a few of these at the one time. Hence statements about the maximally responsive system must always be considered relative to those studied. It is important nonetheless to have some idea of the frequency of occurrence of particular forms of specificity if they are considered of relevance to problems of individual responsiveness. If, for example, SC responders (as assessed using a typical battery of autonomic responses) are extremely rare, an argument that individual response specificity may confound the determination of individual differences in responsiveness in SC lacks force.

The several studies which followed the Laceys' work and sought to replicate and extend it provide support for most of their essential findings and go some of the way towards answering the question of the frequency of different types of responders. The central finding of stereotypy would not seem in doubt as a consequence of the studies of Wenger, Clemens, Coleman, Cullen, and Engel (1961), which provided a close replication of the study of Lacey et al. (1953), and of Engel (1960) and Engel and Bickford (1961), which provided particularly stringent tests of the concept. The study of Wenger et al. (1961) also indicated that stereotypy as expressed by the coefficient of concordance was more likely to be demonstrated using response level measures than change scores; 83% of their sample showed statistically significant stereotypy with the former whereas only 40% showed stereotypy with the latter.

A more recent study by Sersen, Clausen, and Lidsky (1978) further supports the importance of the measurement unit in demonstrating the phenomenon. Their review of studies to that date indicated that stereotypy was most clearly shown when response levels were measured and that "lesser degrees of stereotypy were found using Lacey's (1956) autonomic lability (AL) score" (p.60). They argued for a method of measurement which was a variant of that employed by Engel (1960) and found more evidence of stereotypy with this revised method in a study directed to a comparison of the two.

The Laceys' finding of stability of the individual's hierarchy of responding over time has not been the subject of scrutiny in the way their other findings have. Johnson, Hord, and Lubin (1963) examined stereotypy in a sample of Navy hospital corpsmen in two sessions separated by an interval of 48 hours. Their conclusion was that contrary to the findings of the Laceys the pattern of responsiveness for individuals was not reproducible across sessions. They noted that whereas Lacey had used one condition, cold pressor, to evaluate stereotypy, they employed a number, and that this difference in design may have been responsible for the conflict in findings. While it is of course legitimate to speak of stereotypy with respect to the one stimulus situation, the power of the concept lies in its generalization across a range of stimulus situations. The findings of Johnson et al. (1963) question the stability of the phenomenon when a range of situations are employed, and as such raise doubts about the stability of stereotypy as an individual difference variable.

While not all studies reported since those of the Laceys have included information on the nature of the specificity or stereotypy expressed by individual subjects, several have, and these indicate that SC responding is characteristic of only a small group of subjects. Engel (1960) and Engel and Bickford (1961) found 5.0% of subjects were significant SC responders. Roessler, Greenfield, and Alexander (1964) reported that 11.1% of subjects displayed stereotypy on SC. Crooks and McNulty (1966) reported 16.7% of normals and 30% of schizophrenics demonstrated SR as their significantly most reactive physiological variable. Sersen et al. (1978), however, reported no SC responders in their sample.

Given this relatively low proportion of SC responders in non-clinical samples, it is perhaps not surprising that the question of correlates of this characteristic has not been explored. The only reference to this appears in the report by Roessler et al. (1964). They divided subjects into three equal sized groups on the basis of subjects' scores on the ego strength (Es) scale (Barron, 1956) of the MMPI. They found that fewer subjects with scores in the middle range demonstrated specificity compared to the other two groups. Further, of the subjects demonstrating specificity, idiosyncratic hierarchical patterns were unrelated to Es. As well as this paucity of data on personality correlates of specificity, the more limited question of correlates within the electrodermal system remains unexplored. Although logically distinct from absolute reactivity, it may prove empirically that relative reactivity and absolute reactivity are correlated, with those subjects who are most reactive in SC also showing the largest magnitudes of response in this system. In general, however, specificity has been treated as a separate research topic to that of individual differences in responsiveness.

Application to PDD

To date there has been little systematic investigation of individual differences in PDD. The few studies currently available on the topic have examined the relationship between selected indices of electrodermal activity or personality and differential responsiveness in particular deceptive contexts but have not attempted to deal with the range of indices available. Nor has there been much theory developed about the relationships observed. Indeed, hypotheses are most often based on what their formulators admit are intuititive grounds (see e.g., Waid & Orne, 1980), and there has been little attempt to place predictions in the context of theorizing about PDD in general. As briefly reviewed in Chapter 1, such theorizing involves reference to concepts of attention, arousal, and particularly the orienting response. The present section reviews the studies which have been reported and attempts to develop more theoretically based hypotheses to guide research.

Two studies have examined the relationship between CPI socialization and PDD. Raskin and Hare (1978) examined differences in detectability following a mock crime (theft) in a sample of prisoners classified as psychopathic or non-psychopathic and differing in terms of score on CPI socialization. There was no significant difference in detection rates for psychopaths and non-psychopaths. The total field evaluation scores for subjects with low scores on socialization were, however, significantly greater than those for subjects with high scores i.e., "differentiation between 'guilty' and 'innocent' subjects was better for low socialization compared to high socialization subjects" (p. 131). In a study by Waid, Orne and Wilson (1979) using a sample of undergraduate subjects those guilty of a mock crime but who were classified as "innocent" on the basis of a polygraph examination showed significantly lower scores on CPI socialization than those detected.

Subjects innocent of the mock crime and incorrectly classified as guilty were found to have significantly higher mean scores on socialization (38.5) than innocent subjects correctly detected (31.6). A further observation in this study was that subjects with low scores on socialization showed smaller SCRs during deception.

Bradley and Janisse (1981b) investigated the relation between extraversion and neuroticism, as measured by the EPI, and detectability using SRR following a mock crime. Half the subjects ($\underline{n} = 48$) participated in the mock crime and the other half did not. These researchers failed to find any relation between neuroticism and detectability. However they reported that 81.3% of extraverts (subjects with high scores on the E scale) and 54.2% of introverts (subjects with low scores on the E scale) were correctly detected. Further, an analysis of variance demonstrated that, compared to introverts, innocent extraverts were more likely to be found innocent and that guilty extraverts were more likely to be found guilty. Bradley and Janisse (1981b) interpreted their findings in terms of differences in responsiveness for the two personality types. Innocent extraverts were relatively more responsive to control questions according to Bradley and Janisse, guilty extraverts were relatively more responsive to relevant questions, and guilty introverts were responsive to both relevant and control questions.

A study by Balloun and Holmes (1979) examined the relationship between score on the psychopathic deviate (Pd) scale of the MMPI and detectability using SRR. Of the sample of 34 male subjects 18 were known to have cheated on a written examination and 16 had not. These researchers failed to find any relation between detectability and Pd scores.

Waid and Orne (1980) in two studies using a mock agent test examined the role of electrodermal lability in detectability. In their first study, the number of NSRs of guilty subjects were significantly correlated with the number of code words that were detected ($r = .56$, $p < .01$). This was not the case for innocent subjects ($r = .37$, $p > .05$). Subjects correctly classifed as guilty displayed more NSRs than subjects incorrectly classified as innocent ($p < .025$). In the second study, a comparison was made between innocent subjects who were misclassified on either a CQ test or a GKT and innocent subjects who were correctly classified. False positives displayed more NSRs during rest and during the polygraph test. Thus Waid and Orne's study indicated that subjects showing fewer NSRs who are deceptive are less likely to be detected than deceptive subjects who show a high frequency of NSRs, and that truthful "labiles" are more likely to be found "deceptive" than truthful "stables." Lability would thus seem to be an important predictor of detectability and one according to Waid and Orne (1980; p.4) which is independent of socialization.

Waid, Wilson and Orne (1981) employed a mock agent paradigm to study the effect that electrodermal lability has on the detection of deception when cardiovascular, respiratory and electrodermal measures are employed. These researchers found a significant correlation between lability and detection on a CQ test for guilty subjects ($r = .35$, $p < .025$) but not for innocent subjects ($r = -.05$, $p > .05$). Contrary to Waid and Orne's (1980) finding, Waid, Wilson, and Orne (1981) failed to find a significant effect for lability when using a GKT. A further analysis involved dividing subjects into the upper (labiles) and lower (stables) thirds of the lability continuum. The number of responses to relevant questions that exceeded the responses to adjacent control questions were totalled for guilty and innocent subjects in labile and stable categories. The number of responses so totalled were separately compared on the electrodermal, cardiovascular, and respiratory measures of the CQ test. These researchers found that labile subjects responded to more relevant questions than to adjacent control questions than did stable subjects. This relation held regardless of subjects guilt or innocence, or of the physiological measure employed in detection.

Giesen and Rollison (1980) assumed the existence of a relation between individual response stereotypy and detectability, and then examined the relation between subjects' anxiety and detectability. Stereotypy was defined in terms of score on a self report scale, Stern's Perceived Somatic Reactions Questionnaire (PSRSQ) (Stern & Higgens, 1969). On this questionnaire subjects rank order 11 somatic changes in terms of their awareness of each during stressful situations. Edelman (1972), in demonstrating that subjects in an imaginary fearful situation become reactive on the same physiological index as that which they reported on the PSRSQ, provided some validity for Stern's self report scale. Giesen and Rollison reported that guilty subjects with high scores on the measure of anxiety, the Activity Preference

Questionnaire, (Lykken, Tellegen, & Katzenmeyer, unpublished report) were significantly more responsive to pertinent questions than guilty subjects with low scores on the anxiety measure. The low anxious guilty subjects did not differ significantly in responsiveness from innocent subjects. Further, the only significant difference between responsiveness to pertinent and to neutral items occurred in high anxious guilty subjects. The Giesen and Rollison study cannot be considered a test of the role of specificity in detection. Not only was a self report measure of specificity employed, but the design called for control of the specificity factor rather than for its manipulation.

This survey of the literature on individual differences in PDD indicates that of the three possible dimensions of electrodermal responsiveness only one, lability, has been studied in any detail and even here a number of questions have yet to be answered. For example, it is not clear whether the contribution of lability to detectability reported by Waid and coworkers is independent of any contribution from the other presumptive dimensions. The correlation may be explained more parsimoniously, for example, in terms of a general reactivity factor that influences both detectability and lability. The independence of these two dimensions and their relative contributions to detectability needs to be established. Waid and Orne (1980) did not consider alternative explanations for the lability correlation, although they did note that the contribution of lability is independent of that of the CPI Socialization variable which, at least in one study, was found to relate to amplitude of SCR. But the data on self report correlates of detectability is even more sparse and equivocal. For example, is the correlation between socialization and detectability a function of a correlation with the superfactor, extraversion, since Eysenck (1957) has implicated differences in extraversion in the process of socialization and extraversion has been found to be one of the major factors in analyses of the CPI (Nichols & Schnell, 1963). Clearly the point raised at the conclusion of Chapter 1 concerning the nature of the basic dimensions of electrodermal activity, their relation to more general aspects of personality and to detectability cannot be decided on the basis of the data currently available.

The study undertaken to clarify these points is described in the following chapter. Its execution and the analysis of the data that resulted were guided by a number of hypotheses. The first was that a three factor solution would best represent the dimensionality of electrodermal responsiveness as Podlesny and Raskin implied. The evidence supporting a two factor solution for reactivity and lability indices has been reviewed earlier and the lack of evidence on the position of specificity with respect to the other two has been noted. A three factor solution was thus a working hypothesis. Second, it was expected that any relationships between these dimensions and self report measures of personality could be accounted for in terms of the factors of extraversion and neuroticism. Apart from their wide use in previous studies of correlates of EDA, there is a fair measure of agreement in the psychometric literature (see e.g., Kline, 1979) that these factors can be isolated in most self report tests and are the most general and, though this is more contentious, generalizable factors. The examination of these factors in relation to responsiveness and detectability was therefore considered a starting point, although as the earlier review makes clear correlations between these factors and indices of EDA have not proved robust. As well as these quite broad hypotheses, a number of more specific hypotheses regarding relationships with detectability were sought. An immediate difficulty was encountered, however, in that intuitively plausible hypotheses and ones for which some supporting evidence was available were not consistent with the results of a theoretical analysis of the problem of detectability. It is intuitively plausible to argue that the more responsive a subject electrodermally the more likely he or she is to be detected in that system. (Responsiveness here is used generally to include lability, reactivity, and specificity.) This line of argument implies, however, that differential responsiveness, the basis of detection, is positively related to absolute responsiveness and this is not necessarily the case as a consideration of the psychometrics of difference scores indicates. Difference scores have been discussed by Cronbach and Furby (1970) and Linn and Slinde (1977) among others. Linn and Slinde (1977) provided the following formula for describing a difference score in terms of the variance of the pre- and post-scores, which give rise to the difference score, and the intercorrelation between the pre-score and the difference between a post-score and that pre-score.

$$Pxd = \frac{Pxy\sigma y - \sigma x}{\sqrt{\sigma^2 x + \sigma^2 y - 2Pxy\sigma x\sigma y}}$$

In this formula X is a pre-score and Y is a post-score, D is the difference between Y and X, σx and σy are the standard deviations of X and Y respectively, Pxd is the correlation between X and D, and Pxy is the correlation between X and Y. Inspection of this formula indicates that the correlation between the pre- and difference scores is typically negative. This results because, first, the correlation between X and Y must be less than one and, second, the standard deviations of the pre- and post measures approximate each other in magnitude. Hence $Pxy\sigma y$ will be less than σx and because σx has a negative value the numerator and hence the correlation between the pre- and difference score will be negative. A positive correlation only results when the standard deviation of the post-score is greater than the standard deviation of the pre-score.

Now in a PDD situation response to the control or irrelevant question can be thought of as the pre-score and the difference between the control and critical questions (the subject's differential responsiveness) as the difference or D score. It follows that factors which correlate positively with the prs-score will enter into the same relationship with the D score, which according to the foregoing analysis is negative in sign. A responsive subject, one who shows large amplitude responses to control stimuli, should thus be expected to show a smaller D score, which is of course the reverse of the intuitive prediction. It is interesting to note that the prediction of Balloun and Holmes, that the less responsive extraverts should be more detectable, is consistent with this psychometric analysis although it was not predicted on it.

The conflict between the intuitively plausible and the theoretically deivable is not reduced if the question is looked at in terms of OR theory, and Ben-Shakhar's dichotomization theory in particular, instead of psychometrically. Dichotomization theory predicts that rapid habituation to the neutral stimuli leads to greater differential responsiveness between neutral and target stimuli. Stable subjects, those who habituate rapidly, are thus more likely to demonstrate greater differential responsiveness and hence be more detectable. But this of course is contrary to the observations of Waid and colleagues that labiles, not stables, are more detectable. Waid, however, used NSR frequency rather than habituation speed as the index of lability, though this should not be significant if in fact lability includes both NSR and habituation components.

In view of this conflict it was not possible to formulate directional hypotheses. All that could be asserted was that on the basis of previous findings or from a theoretical analysis of the problem, a correlation would be observed.

In summary, the study sought to establish the interrelations of measures of reactivity, lability, and specificity within the electrodermal system, the separate and joint contributions of these measures to individual differences in detectability in that system, and the relationships of these measures and of detectability to the major self report dimensions of personality, extraversion and neuroticism.

CHAPTER 3
METHOD

Overview

Testing the hypotheses formulated in the closing section of the previous chapter required the assessment of individual differences in electrodermal activity and the assessment of detectability for a sizable sample of subjects. To provide a statement of some generality it was considered advisable that the assessment of individual differences in electrodermal activity be conducted separately from the determination of detectability. Ideally, testing in separate sessions would have ensured that any relationships observed were not a function of the state of the subject at the time of testing. Given the large numbers involved this was not considered feasible. Instead a procedure was designed in which the assessment of individual differences in responsiveness preceded the tests for detectability.

The first phase of the study involved the subject completing four tasks while SC, HR, and RR (respiration rate) were monitored. The tasks were selected (a) to provide a range of arousal so that reactivity and specificity could be assessed, and (b) to include tasks typically used in the assessment of lability. In the case of the latter the tasks were relaxing without any response demands and presentation of a series of tones. In the case of the former the tasks were listening to a count-up and performing mental arithmetic. The count-up was a procedure first used by Deane and Zeaman (1958) and more recently by Hare (1965). The subject hears a count from 1 up and on a designated number in the series a loud noise or shock is presented or is threatened to be presented. The count-up task has been used by Epstein (e.g., Epstein & Clarke, 1970; Epstein & Roupenian, 1970) to study anticipatory electrodermal arousal. Mental arithmetic in which the subject is asked to calculate the answers to arithmetic problems often under time pressure has been widely used in the psychophysiological laboratory to induce increased arousal (e.g., Lacey & Lacey, 1958; Sersen et al., 1978). The low arousal tasks, relaxation and tone series, used to assess lability are those typically employed for this purpose (e.g., Crider & Lunn, 1971).

The selection of response systems was dictated in part by availability and in part by a consideration of those used in laboratory studies of PDD. SC was of course the focus of concern in the study but other measures were needed to assess degree of specificity in this system. RR has been widely used in both laboratory and field studies (e.g., Raskin & Hare, 1978; Reid & Inbau, 1977) and was included for this reason. The third system included, HR has been used in laboratory studies of PDD. In the field situation relative blood pressure and not heart rate is employed, but apparatus for continuous recording of this response was not available.

For the determination of individual differences in detectability, it was considered appropriate to use more than one test procedure and more than one method of questioning to ensure some generality in the results. The card test was selected because of its wide use in laboratory studies, and this was supplemented by a mock agent test. In the case of the latter both the CQ test and GKT techniques of interrogation were employer. Separate samples of subjects were used for the card test and the mock agent, principally because of time constraints. The volunteer population in which the study was conducted could not be expected to remain for longer than the hour and a half for which the session was scheduled.

Two variations were included in the card test, again in the interests of increasing generality. The first was to question all subjects under each of three response sets: having the subject reply no to each question, having the subject reply yes to each question, and having the subject not make any verbal response. Typically the card test requires subjects to answer "no" to each question. Gustafson and Orne (1965) reported that the frequency of correct detections was significantly greater under the requirement to say "no" than if no response or an irrelevant response was required of the subject. Kugelmass, Lieblich and Bergman (1967) reported, however, that frequency of correct detections was not significantly greater when subjects were required to respond "yes" to all questions than when they were required to respond "no". Individual differences in responsiveness may be less important under the no set as the need to lie verbally may increase responding for all subjects.

The other variation to the card test was to include a group of subjects who were questioned about cards they did not select. The purpose of this group was to act as a control to establish the role of card selection in any results obtained. When the measure of detectability and the estimate of responsiveness which is to serve as the predictor are both drawn from the same psychophysiological response system the question arises whether any positive correlation observed between the two is tautological. That is, the correlation may have nothing necessarily to do with deceptiveness but arise simply because of the commonality in the method of deriving both variables. If it can be shown, however, that the correlation exists only under conditions in which subjects are deceptive then a much more precise meaning can be attached to it.

In the card test every effort was made to ensure that the subject's status was not known until all scoring of the data was complete. With only one experimenter a double-bind procedure was not possible, but procedures described in detail below were adopted to prevent bias in data reduction. In the mock agent test the problem of bias is more significant since in this test all subjects were guilty and the experimenter knew this. As Podlesny and Raskin (1977) noted, "If the examiner displays a bias towards a particular outcome or treats the subject in an accusatory manner, this could produce reactions that may contribute to an erroneous outcome" (p.788). Although subjects in the present study were not treated in an accusatory manner, the examiner did know that all were guilty and this knowledge could have influenced the outcome. The extent to which bias of this sort influenced the results of the mock agent test, it must be admitted, is unknown.

One other factor considered in developing procedures for the study was the degree of subject motivation to be induced. Studies of PDD have shown a positive relationship between level of subject motivation and accuracy of detection. Gustafson and Orne (1963) demonstrated that accuracy rates were substantially higher when subjects were motivated to deceive as a consequence of ego involving instructions and provision of a monetary incentive. In that study, two groups of subjects selected one of five cards. One group, the motivated group, was informed that they were going to be tested to see how effectively they could conceal information and that while this was a very difficult task, they could succeed if they were of superior intelligence and possessed exceptional emotional control. Further, if they succeeded they received a $1 bonus. The subjects in the non-motivated group did not receive these instructions nor any bonus. Gustafson and Orne (1963) found that subjects in the motivated group were detected significantly more often than subjects in the non-motivated group ($p < .05$), and that detection rates were at a much higher than chance level, which was not the case for subjects in the non-motivated group. Work from Raskin's laboratory has also emphasised the importance of motivating subjects prior to conducting tests of deception. For example, Raskin and Hare (1978) motivated prisoners in a mock theft paradigm by providing a $20 bonus for successful performances. When an inconclusive category was excluded these researchers obtained a 96% detection rate.

It was decided not to attempt to induce high levels of motivation in the present study because concern focused on the role of individual difference factors in PDD. In the light of studies just reviewed, strong motivational press could be expected to reduce variation in detectability among subjects and reduce the opportunity to show correlations with personality factors. If, for example, in the extreme case all subjects were detectable, no correlation with the factors of concern in the present study would be possible. To avoid ceiling effects, therefore, it was decided not to attempt to induce high levels of ego involvement or to offer a monetary incentive, even though this represented a departure from the optimal conditions for PDD.

In summary, two samples of subjects were employed. Each subject was tested in a single session which consisted of three phases: the assessment of individual differences in physiological responsiveness; the assessment of detectability; and the administration of personality inventories. The major procedural difference occurred in the second phase which for Sample 1 ($\underline{n}$ = 126) involved a card test and for Sample 2 ($\underline{n}$ = 84) involved a mock agent procedure.

Subjects and Design

Sample 1. A total of 63 male and 63 female undergraduates enrolled in the introductory psychology course at the University of New England were included in Sample 1. Two thirds of the sample (42 males and 42 females) constituted the experimental group. The remaining third of the sample (21 males and 21 females) constituted the control group. Due to equipment malfunction 5 subjects were lost from the analysis.

For the assessment of individual differences in responsiveness, subjects were studied under four conditions: relaxation, tone series, count-up, and mental arithmetic. All subjects completed the relaxation condition first. The order in which the remaining three conditions was completed was counterbalanced across subjects. With three conditions there are three possible orders and six possible sequences. Equal numbers of subjects of each sex were allocated to each of the six sequences.

For the assessment of detectability, all subjects completed three conditions: a yes condition in which subjects responded "Yes" each time they were asked whether they had seen a certain card, a no condition in which subjects responded "No", and a mute condition in which subjects made no verbal or physical response when presented with each card. Counterbalancing was again employed, with equal numbers of subjects being assigned to each of the six possible sequences.

Sample 2. The source of this sample of 42 males and 42 females was the same as that for Sample 1 and the same design was followed with this sample for the first phase of the study.

For the assessment of detectability, four conditions were employed: two relevant-control tests, and two guilty knowledge tests (GKT). In one GKT the subject was required to respond "No" to all questions and in the other (a mute condition) no response was required. The relevant-control tests preceded the yes condition of the GKT which in turn preceded the mute version of the GKT in all cases. This is the order employed in other studies (e.g., Waid & Orne, 1980; Waid, Orne, & Wilson, 1979) that have employed both methods of interrogation.

Apparatus and Materials

Physiological activity was recorded on a four-channel Grass polygraph, model 79D. One channel was used to record SC directly using the circuit proposed by Venables and Christie (1973, p.102) which provided a constant voltage of .5V across the electrodes. Bipolar recordings of SC were made using Ag/AgCl electrodes of 12mm diameter affixed, with KY jelly as the electrolyte, to the first and third fingers of the left hand. A second channel was used to record respiration from a strain guage (Phipps and Bird pneumograph) positioned around the subject's chest and connected to a Grass volumetric pressure transducer (PT5A) and in turn to a 7P1 low level DC pre-amplifier. A third channel was used to record EKG (electrocardiograph) from 32mm silver electrodes attached to the subject's wrists. The signal was fed to a Grass low-level DC preamplifier, model 7P1.

All instructions and stimuli excluding the white noise burst used in the count-up condition were recorded on a Sony TC 105 reel-to-reel tape recorder and presented to the subject through Akai stereo-headphones, model ASE-8S. The experimenter monitored the tape through a second pair of headphones. The white noise burst was produced by a locally built white noise generator, and fed through the headphones. Its duration was controlled by a locally built electronic timer and was manually scheduled by the experimenter depressing a key.

Two pencil and paper personality questionnaires were employed in the study. In Sample 1, 84 subjects completed the Eysenck Personality Questionnaire (EPQ) (Eysenck & Eysenck, 1975) and a short form, containing 32 items, of the socialization scale of the CPI. All subjects in Sample 2 completed the EPQ and the full socialization scale of the CPI.

Procedure: Phase 1

The psychophysiological laboratory consisted of two adjoining rooms, one for the subject and one which housed the stimulus equipment and recording apparatus. On arrival at the laboratory subjects were requested to wash their hands and were then seated in a reclining chair. Rings and watches were removed and electrode sites on the insides of the wrists prepared by brisk rubbing with methylated spirit. The EKG electrodes were attached using elasticized bands. The electrolyte was commercial EKG paste which was applied to both the skin surface and electrode. Skin conductance electrodes were taped to the medial phalanges of the ring and index fingers of the left hand. The electrolyte was KY jelly which was carefully applied so as not to spill out from under the lip of the electrode. The strain guage was placed around the subject's chest to monitor respiration.

When subjects were recruited for the study they were told that the experimenter was interested in the responses of the autonomic nervous system and that sensors would be attached to subjects so that skin conductance, heart rate, and respiration could be monitored. This general information was repeated at the commencement of the recording session with advice that electric shock would not be used at any stage and a request that the subject minimize movement. Specific instructions for each phase of the experiment, the subject was informed, would be presented through the headphones when appropriate. These were then put in place, and the experimenter retired to the recording room.

All subjects completed the relaxation condition first. The instructions which introduced it were as follows:

> "For this part of the experiment I would like you to relax as much as possible. Nothing unpleasant is going to happen to you. Make yourself comfortable in the chair. Let your whole body go relaxed and lazy. Let your arms and legs go loose with your hands on your lap and your legs stretched out comfortably. I want you to feel very comfortable, lazy and very relaxed. I will tell you when this phase of the experiment is completed."

At the first and third minute the following instruction was presented.

> "Still very lazy, feeling comfortable, very relaxed, that's good."

To terminate this as all conditions in the first phase the experimenter said,"O.K. Thanks very much."

Subjects completed the remaining conditions in the order to which they had been assigned according to the counterbalanced design. The specific instructions for each condition follow. For mental arithmetic they were:

> "In this phase of the study I want you to do some

mental arithmetic. Your main task is to count backwards from the number 1000 by sevens. Begin with the number 1000, take seven from it, take seven from the remainder and so on in this way, 1000, 993, 986, etc. Continue counting backwards until you reach zero, then begin the count again. This task is to be done silently. From time to time I will give you arithmetic problems to solve. When you are given a problem, cease counting, work the solution in your head, as quickly as you can, and then report the answer to me. Once you have completed the problem, resume counting backwards from where you left off until another problem is presented. Then cease counting, work the problem in your head, report the answer to me and resume counting. You will be told when this phase of the study is completed. Remember, count backwards from one thousand silently but report the answers to the specific problems out loud so that I can record them. Work quickly please and remember do not move your hands. Begin."

At the first minute the experimenter said,

> "Multiply 17x13. Report your answer to me as quickly as you can, then resume counting."

The same verbal instruction was given at the second, third, fourth, and fifth minute but the figures used were 14x19, 16x17, 13x18, and 15x17 respectively.

For the tone series, the instructions were

> "In this phase of the study you will be presented with a number of soft tones and I just want you to sit and listen for each of the tones."

There were 10 1000Hz tones, 60dB in intensity and .5 sec. in duration which were presented at 30 sec. intervals.

Instructions for the count-up were:

> "In this phase of the study you will hear someone counting from 1 to 15. On the count of 10 you will hear a loud noise. I just want you to sit and listen to the person counting. There will be a loud noise on the count of 10".

The interstimulus interval for the count was 15 sec., and the intensity of the noise burst presented at count 10 was 100dB.

Each condition took approximately 6 mins. to administer and was followed by a short break.

Procedure: Phase 2

The procedure followed in the second phase of the study, the assessment of detectability, differed for subjects in Sample 1 and Sample 2. For subjects in Sample 1, the procedure for administering the card test was complicated by the decision to include a control group in the design, i.e., subjects who were questioned about cards they had not selected. This decision was implemented with a sub-sample of 84 subjects in Sample 1, half being assigned to the experimental and half to the control group. To ensure that the experimenter remained blind as to the guilt or otherwise of the subject being tested, and yet maintain the counterbalancing and have equal numbers of males and females in the two groups, the following procedure was adopted. The six cards, the 2,3,5,8,9 and 10 of diamonds, to be presented to subjects in the experimental group were placed in an envelope which was placed in a larger envelope. The ace, king, and queen of clubs and of spades were also placed in a small envelope which in turn was placed in a larger envelope identical to that holding the cards for guilty subjects. The two larger envelopes, one containing the envelope of cards for the experimental group and one containing the envelope of cards for the control group, were shuffled for each subject. The experimenter selected one envelope, opened it and noted, on the smaller envelope contained inside, the subject's sex and the order in which the verbal response conditions were to be presented. Selection at random was maintained until late in the series when all subjects required for one of the groups had been tested, as indicated by the running total on the smaller envelope. The remaining subjects were assigned to the other group. Which group had been completed first remained unknown to the experimenter.

Once a card set had been determined for the subject the experimenter entered the subject's room, laid the six playing cards face down on a chair, and said,

"Here are six playing cards. When I leave the room I would like you to turn one card over and write the value of the card on this piece of paper and seal the piece of paper in this envelope. Then turn the card you selected face down so that I do not know which card you selected."

The experimenter then left the room and returned when the subject's task was completed. The experimenter collected the cards taking care to replace them in the smaller envelope so as not to learn inadvertently the card chosen or the group to which the subject was assigned. The experimenter then explained that three tasks would follow, each task being preceded by a separate set of recorded instructions. Instructions for the yes condition were:

"In this phase of the study I am going to ask you which card you chose. I want you to answer "Yes" to every question. That is, even if you did or did not choose the card that I state, answer "Yes" out loud. I will present the card numbers shortly."

The first question was,

"Did you choose the ace of diamonds?"

This was to act as a buffer item prior to the presentation of the six cards. Each card was presented twice, in the following order: 3, 8, 5, 10, 9, 2, 5, 9, 10, 2, 8, 3 with each card taking the place of the ace in the above question. Questions were presented at the rate of 1 per 15 sec. Instructions for the no condition were:

"In this phase of the study I am going to ask you which card you chose. I want you to answer "No" to every question. That is, even if you did or did not choose the card that I state, answer No out loud. I will present the card numbers shortly."

Instructions for the mute condition were:

"In this phase of the study I am going to ask you which card you chose. As the numbers are presented to you I do not want you to make any verbal response at all. That is, even if you did or did not choose the card that I state, do not make any attempt to answer. I will present the card numbers shortly."

Rate and order of presentation in the yes and mute conditions were the same as those in the no condition.

At the conclusion of the recording session, the electrodes, strain guage, and head set were removed and the subject escorted into the adjoining room where the personality questionnaires were administered and any questions about the study answered. Subjects were asked not to discuss what happened to them with their friends. The envelopes with the subjects' records of the cards they had chosen were not opened until all scoring of the data was completed.

The procedure followed with subjects in Sample 2 involved the induction of the set for role-playing an espionage agent having in his or her possession secret code words, and then the interrogation of the agent. In the first stage, the induction of the set, the subject was seated on one side of a table, in a room separate from the psychophysiological equipment. On the table there were three sets of cards labeled 'A', 'B', and 'C'. All cards were face down. The experimenter sat on the opposite side of the table behind a partition which prevented any view of the cards. The following was read to each subject:

> "A recent spy case in America was solved using a lie detector. So, we are studying the effectiveness of the polygraph as a lie detector. You are to play the role of an espionage agent who may or may not know certain critical code words. I am going to ask you to select one card from each of the three piles of cards in front of you, one from pile "A", one from pile "B", and one from pile "C". If the cards are blank you are an innocent person who will not know the critical code words.[1] If the cards have words on them, you possess critical code words that only you know. It is important that you do not reveal these words to anyone. If you know the critical code words I would like you to write them down on this piece of paper and place the piece of paper in this envelope. Put the

---

1. There were no blank cards in the sets presented to subjects. Reference was made to blank cards to give subjects the impression that both innocent and guilty subjects were participating. All subjects were, however, guilty, and this may have produced examiner bias as discussed on page 47.

cards face down in their original positions. If you do not know the critical code words, i.e., if the cards that you chose were blank, I would like you to write the word 'blank' three times on the piece of paper and place the paper in the envelope. Could you do that now please...?

On a separate piece of paper I would like you to write down each word 10 times. If you do not know the code words, i.e., if your cards were blank, I would like you to write down your first name 30 times. Turn that piece of paper over. Next, if you know the code words I would like you to put them in reverse alphabetical order. If you do not know the code words, write your surname six times.

Due to a security leak, I am in possession of a pool of words and it is my task to attempt to determine, firstly, if you are in possession of the code words, and secondly which words in the pool of possible code words are critical words that you may or may not know. Your task is to try to convince me that you are innocent and to convince me that none of the words that will be presented to you have any meaning for you. You will be able to do this by not responding physiologically whilst being examined. If you are in possession of the code words you will find it difficult not to respond. Recent American studies have shown accuracy rates in lie detection to be as high as 95% which has led some American authorities to claim that the polygraph is virtually infallible in lie detection. If you wish to save a good deal of time you can confess now and tell me the code words that you are possessing. Do you wish to tell me the code words?[1] The actual interrogation will be very similar to a real life polygraph interrogation."

The experimenter and subject then walked a short distance to the

---

1. No subject opted to reveal the code words.

psychophysiological laboratory where the following general instructions were provided:

> The interrogation will consist of two sections and will be very similar to a real life polygraph examination. That is, the questions to be asked in section one will be reviewed with you to eliminate any surprise, ambiguity, or uncertainty which might interfere with the interpretation of the polygraph record. In section one of the interrogation you will be asked the following questions in this same order. I would like you to answer "yes" or "no" as I read them to you and to answer the same way during the interrogation.
>
> The questions[1] are:
>
> 1. Are you in Armidale, New South Wales now?
> 2. Do you own a car?
> 3. Are you playing the role of a courier carrying critical code words?
> 4. Do you ever smoke cigarettes?
> 5. Have you been given critical code words that only you know?
> 6. In your whole life have you ever stolen anything?
> 7. Have you any assignments due this week?
> 8. In your whole school career did you ever do

---

1. These questions were formulated using examples provided in a number of papers (e.g., Waid, Orne & Wilson, 1979). One difficulty with them which must be acknowledged, is that the control questions are all of the type labelled by Podlesny and Raskin (1978) "non-exclusive". That is, they are questions which lack any specific time of reference and as such are to be distinguished from "exclusive" questions which are specific. Podlesny and Raskin (1978) reported that the latter are more effective in identifying guilty subjects.

anything that was considered wrong?

9. Were you recently given the name of a type of tree, type of bird and a colour?"

The questions were repeated and it was explained that on the second presentation proper two additional questions would be added.

> "They will be: 'Do you intend to tell anyone what this experiment entails?' 'Did anyone tell you what this experiment entailed?'"

If subjects required clarification on any point or if they answered "yes" to a control question, the experimenter again explained the situation to subjects before proceeding to the next question. For example, if a subject responded in the affirmative to the question "In your whole life have you ever stolen anything?", the examiner asked for details concerning the theft. Invariably petty theft was involved. The examiner asked whether there was any other incidents of theft. When the subject answered "no" the examiner said, "When you are asked the question 'In your whole life have you ever stolen anything?' I want you to think about that but exclude the incident of theft that we have just discussed."

> "In the second and final section of the interrogation you will be asked if any of 21 words have special meaning for you as critical code words. The 21 words will be presented on two occasions. On the first occasion I want you to answer "no" to every question. On the second occasion I do not want you to make any verbal response. I would like you to sit quietly and listen to the questions. Please do not go to sleep. Throughout the experiment try not to move about."

The experimenter explained that instructions would precede all conditions and that the instructions just given to subjects did not have to be strictly remembered, as they would be repeated through the headphones. Headphones were then placed on the subject so that tape-recorded instructions could be heard. Following a 20-sec. pause the nine questions listed above were presented. Following another 20-sec. pause they were presented again with the two additional questions. A pause between 14 and 16 sec. separated each question. On completion of the relevant-control test the experimenter said,"OK, this completes the first section of the interogation."

The next stage, the GKT, commenced with the following instructions:

> "In this the second section of the interrogation you will be asked if any of 21 words have special meaning for you as critical code words. The 21 words will be presented on two occasions. On the first presentation I want you to answer "No" to every question. On the second presentation I do not want you to make any verbal response, just sit quietly and listen to the questions. This is now the first presentation. Please answer "No" to every question."

After a 20-sec. pause questions were presented at intervals of 14 to 16 sec. Questions took the form:

"Does the word ... have special meaning as a critical code word?"

Shortly after, instructions for the "mute" condition were presented.

> "This is the second presentation of critical code words. During this presentation I do not want you to make any verbal response. I would like you to sit quietly and listen to the questions."

Following a 20-sec. pause questions were presented between every 14 to 16 sec. in the same form as the first presentation.

The first word in each set of words was a buffer item. The order of presentation of each group of words was as follows: eagle, crow, pigeon, owl, wren, robin, sparrow, spruce, oak, mulberry, gum, elm, pine, poplar, marone, purple, orange, brown, grey, green, blue.

At the conclusion of the mute version of the GKT and prior to requesting subjects to complete the personality questionnaire, the experimenter asked subjects if any of the words had any meaning in a context external to the experiment. A few subjects replied in the affirmative in which case the word with "special" meaning was noted on the front of the subject's record so that reference could be made to the subject's response to the word when records were being scored. None of the words in question were found subsequently to have elicited a larger SC response than any of the other words in its group. Hence the "special" meaning of a certain word did not affect the scoring of responses. As with Sample 1, questions about the study were answered and subjects asked not to discuss it with their friends.

Procedure: Phase 3

The strain guage and electrodes were removed from the subject and the experimenter and the subject went into the adjoining room. The subject was handed several sheets of paper and was requested to answer the questions on them. Subjects were not told that the questions were the basis of a personality questionnaire. When subjects had completed this task they were requested not to give their friends any information regarding the experiment, thanked for their participation, and then left the building. A total of 42 in Sample 1 completed the EPQ and a short form of the socialization scale of the CPI. All subjects in Sample 2 completed the EPQ and all items of the socialization scale.

Data Reduction

For each polygraph record, a number of 5-sec. measurement intervals were identified. For the first phase of the study these were, for the relaxation condition, the periods following the completion of initial relaxation instructions, prior to and following both reminders to continue relaxing, and preceding the onset of termination of relaxation. For the tone series, the measurement intervals were prior to and immediately following each tone. For the count-up they were the periods prior to and immediately following each count. As the noise burst accompanying count 10 evoked startle and consequent disturbance to recording in many cases, it was decided not to measure responsiveness to this count. For mental arithmetic the measurement intervals were prior to and following each problem. Measurement intervals contaminated by movement or respiratory artifacts were discarded.

For each physiological index, measurements of level and amplitude were made using the intervals identified. For SC, amplitude was measured as the largest change in conductance in micromhos occurring in a post-stimulus interval. Level was the SC level at the end of a pre-stimulus interval. For HR, the amplitude measure was the absolute difference in beats per minute between average HR in the pre- and post-stimulus intervals. Level was the average of the 5-sec. pre-stimulus interval. For RR, the measurement intervals were used to define the commencing points for scoring, rather than to define only the sections of record to be scored. The commencing point for the pre-stimulus measurement interval was the first peak or trough in the respiratory cycle following stimulus offset. For the pre-stimulus measurement interval it was the last peak or trough immediately preceding stimulus onset. A rate measurement was obtained by determining the time taken for one or two complete cycles forward (for the post-stimulus) or back (for the pre-stimulus) from the commencing point, and converting this to a cycles per minute index. One or two cycles was used depending on the inter-stimulus interval in the condition for which measurements were derived. Where the inter-stimulus interval was relatively long (relaxation, tone, and mental arithmetic conditions) two cycles were measured, unless artifact was present necessitating the rejection of a cycle and its replacement with a subsequent one. Where the inter-stimulus interval was short (the count-up, the card test, and the mock agent test), one cycle only was measured. RR amplitude was then calculated as the difference between the pre- and post-stimulus rates, and level as the pre-stimulus rate.

The measures of SC used are those typically employed in the psychophysiological literature. The HR and RR measures, however, require comment. HR unlike SC is a bidirectional response, with both increases and decreases in HR being possible. The exact pattern of HR change in response to stimulation has been the subject of considerable investigation (see e.g., Siddle & Turpin, 1980). A number of stimulus and organismic factors are involved ranging from the intensity of the stimulus to the age of the subject, and there is, as a consequence, no standard index of HR for use in the individual case. Different researchers select for analysis different features of the profile of HR change following stimulation using considerations, such as the salient features of the profile shown in group data or the expected pattern of change, to guide their choice (Siddle & Turpin, 1980). For present purposes the direction of change was not as significant as the magnitude of change since HR (and RR) were measured principally to assess the comparative reactivity of the electrodermal system. An average measure was judged to provide the necessary information although it was recognized that identification of the maximum or minimum rate following stumulus onset would have provided a more sensitive index. Determination of maxima and minima would have necessitated a beat by beat analysis which was not practicable given the method of recording available and the number of subjects in the study. For subjects changing predominantly in one direction or the other the average measure is meaningful but conservative. It would, however, be a particularly insensitive index where a subject increased and decreased HR by equivalent amounts during the measurement period.

RR is also bidirectional and as such an average value shares the same problems as those of an average HR measure. A further difficulty is that, while HR (or its reciprocal heart period) is widely accepted as a basic measure of cardiac function, rate is but one of a number of competing measures of respiratory activity (e.g., amplitude, inspiration/expiration ratio). Rate was used here for two reasons. The first was that Stein and Luparello (1967) recommended a rate measure in preference to an amplitude measure where, as in the present study, a girth technique of recording is employed. The second was that rate was the index used in previous studies of IRS which had included assessment of respiratory activity (Engel, 1960; Engel & Bickford, 1961; Johnson et al., 1963; Sersen et al., 1978). As respiration was included in the study primarily for the assessment of IRS it was considered most appropriate to follow past practice in this regard.

Scoring for phase two, the assessment of detectability, varied with the procedure employed. In Sample 1 measurement intervals coincided with the presentation of cards, excluding the buffer items. The SC, HR, and RR measures were the same as phase one. Measurement intervals for Sample 2 coincided with relevant and control questions in the relevant-control test and with the presentation of the critical stimuli (the birds, trees, and colours) in the GKT. SC measurements were the same as employed in Sample 1. The measurement of HR amplitude was the increase in post-stimulus beats per minute over pre-stimulus beats per minute determined over a 5-sec. period. The measurement of RR amplitude was the increase in post-stimulus cycles per minute over pre-stimulus cycles per minute determined in the same manner as in phase one.

CHAPTER 4
ASSESSMENT OF INDIVIDUAL DIFFERENCES

The present chapter reviews the results for the first phase of the study, the assessment of individual differences in physiological responsiveness, and examines the relationship of these measures to the self report measures of personality. In all 205 subjects participated in the first phase, and the data for these subjects are summarized below.

Reactivity and Lability Measures

A mean SCR amplitude and a mean SC level score were derived for relaxation, tone series, count-up, and mental arithmetic conditions, as the averages over amplitude and level measures in each condition. For example, in mental arithmetic there were five amplitude measures per subject. These were averaged for each subject to give an amplitude score for mental arithmetic. Similarly, the five level indices were averaged to provide one level score for mental arithmetic. As well as amplitude and level scores, a lability index was derived for each condition. In the relaxation and mental arithmetic conditions the lability score for each subject was the total number of NSRs. In the tone series and count-up conditions the lability score was the total number of SC responses evoked by the tones and numbers.

The scores for each subject on each of these reactivity and lability indices appear as part of Appendix A. The means and standard deviations are presented in Table 4.1 with the intercorrelations among them. (All data analysis reported in this chapter were performed using SPSS (Statistical Package for the Social Sciences) version 8 on the University of New England's DEC System-20.) One way analyses of variance with repeated measures were conducted on the mean amplitude and level scores. The lability scores across the four conditions were not comparable because of differences in the number of stimuli or the time over which the scores were derived, and hence were not subjected to analysis of variance. For amplitude, the comparison of conditions yielded an F of 263.73 ($df = 3, 811$; $p < .001$). Comparison of the means for the four conditions using the Newman-Keuls method (Winer, 1962) indicated statistically significant differences ($p < .01$) between all means. As inspection of Table 4.1 indicates, amplitude was largest for mental arithmetic and smallest to the tone stimuli. Relaxation might have been expected to evoke the smallest response, but the stimuli to which amplitude was measured in this condition was the experimenter's voice (see p. 60) which was heard after a period of silence. The nature and surprisingness of this stimulus no doubt accounts for the relatively greater amplitude in this condition. For the level measure, the F for conditions was 89.98 ($df = 3, 811$; $p < .001$). Again comparisons of means were performed using the Newman-Keuls method. Results indicated significant differences ($p < .01$) between relaxation and both the count-up and mental arithmetic conditions and between the tone series and mental arithmetic conditions, but no difference between the relaxation and tone series or between the count-up and mental arithmetic. Thus SC amplitude and to a lesser extent SC level differentiated among the four conditions.

The intercorrelations among the amplitude scores shown in Table 4.1 can be thought of as indices of the reliability of these scores. They reflect the degree of relationship between the score and itself across different measurements. This is reliability conceived of as consistency of measurement rather than as stability of measurement which depends on the intercorrelation across occasions rather than conditions of measurement. Conceived of in this way, the intercorrations can be used to derive a reliability coefficient, coefficient alpha (Nunally, 1967). The value computed was .71, which is high by psychometric standards (Nunally, 1967). The alpha coefficient for SC level derived in the same way as the amplitude measure was .96.

The average intercorrelation of the response frequency indices was somewhat lower than was the case for the amplitude and level measures. This may be due to the use of two somewhat different definitions of lability, one based on the number of non-specific responses (in the relaxation and mental arithmetic conditions) and the other on the number of stimulus-evoked responses (tone series and count-up). The

Table 4.1

Means, Standard Deviations and Intercorrelations for

Reactivity and Lability Scores derived from Phase 1 of the Study

	Skin Conductance Amplitude Indices				Skin Conductance Base Level Indices				Lability Indices			
Variables	1	2	3	4	5	6	7	8	9	10	11	12
Means	.54	.17	.34	1.85	8.20	9.47	11.39	12.31	11.55	3.12	4.21	5.88
SDs	.78	.30	.39	1.44	5.77	8.46	8.33	9.21	14.20	2.94	2.75	4.54
Intercorrelations[1]												
1. Amplitude-Relaxation		.75	.68	.60	.62	.60	.64	.57	.55	.46	.39	.34
2. Amplitude-Tone Series			.55	.45	.50	.52	.55	.47	.51	.63	.38	.30
3. Amplitude-Count-up				.58	.54	.52	.65	.53	.46	.38	.60	.43
4. Amplitude-Mental Arithmetic					.55	.55	.61	.62	.31	.29	.29	.34
5. Base Level-Relaxation						.91	.91	.91	.51	.38	.39	.37
6. Base Level-Tone Series							.91	.92	.44	.33	.34	.33
7. Base Level-Count-up								.92	.44	.35	.43	.37
8. Base Level-Mental Arithmetic									.38	.26	.33	.38
9. Lability-Relaxation										.63	.50	.54
10. Lability-Tone Series											.55	.49
11. Lability-Count-up												.52
12. Lability-Mental Arithmetic												

Note: 1. Correlation coefficients in this table are significant at the .001 level.

intercorrelation of relaxation NSRs and frequency of response to tones (.63), however, agrees quite closely with values reported in the literature and on which Crider and Lunn based their argument for a common underlying dimension for the two measures.

In order to summarize the intercorrelations of the amplitude, level, and response frequency indices shown in Table 4.1, a principal components analysis of these data was conducted. The principal components analysis (method PA1 in SPSS) yielded only two components with eigen values greater than unity, and these were rotated according to Varimax. The results of the analysis are summarized in Tables 4.2 and 4.3. Inspection of Table 4.2 indicates that Factor 1 is defined by SC base level and loads the amplitude measures as well, whereas Factor 2 is defined by the response frequency indices and also loads the amplitude measures.

Individual Response Specificity

Data were available on level and amplitude of response to stimulation under each of the four conditions for HR, RR, and SC. From these data an index was derived for each subject of the extent to which skin conductance was the most reactive system. In deriving this index, attention was paid to the large literature on the quantification of physiological responses (e.g., Benjamin, 1963; Heath & Oken, 1965; Lacey, 1956), and a number of preliminary analyses of subsets of the data were conducted using different measurement units. These ranged from uncorrected difference scores to the relatively sophisticated unit proposed by Sersen, Clausen, and Lidsky (1978). The final choice of method was made on the grounds of the most direct procedure which yielded the highest incidence of specificity, irrespective of the system in which it was demonstrated (i.e., the method which yielded statistically significant coefficients of concordance among response systems most often).

Table 4.2

Varimax Rotated Factor Matrix of Reactivity and Lability Variables

		Variable	Factor 1	Factor 2
SC Amplitude Indices	1.	Relaxation	.54	.57
	2.	Tone Series	.40	.61
	3.	Count-up	.48	.58
	4.	Mental Arithmetic	.57	.34
SC Level	5.	Relaxation	.88	.31
	6.	Tone Series	.90	.25
	7.	Count-up	.91	.32
	8.	Mental Arithmetic	.95	.19
Lability Indices	9.	Relaxation	.25	.71
	10.	Tone Series	.11	.79
	11.	Count-up	.20	.65
	12.	Mental Arithmetic	.21	.58

Table 4.3

Eigen Values and the Proportion of Variance Accounted for by Reactivity and Lability Variables in the Factor Analysis

	Variable	Eigen Value	Proportion of Variance
SC Amplitude Indices	1.	6.77	56.4
	2.	1.66	13.8
	3.	.89	7.4
	4.	.76	6.3
SC Base Level Indices	5.	.56	4.7
	6.	.42	3.5
	7.	.35	2.9
	8.	.21	1.8
Lability Indices	9.	.17	1.4
	10.	.08	.7
	11.	.07	.6
	12.	.06	.5

The method involved the use of standardized but uncorrected measures of level.[1] The resulting index is similar to that termed "autonomic tension score" by Lacey and Lacey (1962), the one for which specificity was also most frequently expressed in his studies. The index used here is, however, more conservative in that it does not involve the maximum level reached during a condition but the average level. Like Lacey's ATS it is not corrected for any correlation with prestimulus level. Apart from the observation that such an index yields a higher incidence of responce specificity, it can be justified on two grounds. First, in the situation in which subjects are undergoing a variety of conditions sequentially, the concept of a prestimulus level looses much of its meaning. The prestimulus level for condition B (where A and B are presented sequentially and at short intervals) is the final level for condition A. That is, carry-over from one condition to another in a repeated measures design can vitiate the concept of a prestimulus level. Second, removal of the effect of a prestimulus level, assuming that it could be accurately assessed if sufficient time between conditions were allowed for carry-over effects to decrease to zero, is not appropriate if differences in prestimulus level are characteristic of subjects and the concern is with individual differences. This argument was put by Heath and Oken (1965) in their review of various methods for "undoing" the LIV. They argued that only where differences in prestimulus level were extraneous to the concerns of the study (e.g., temporary, state dependent effects in research on individual differences) should such differences be removed. As the interest of the present study focused on individual differences, removing variance due to individual differences was considered unsound.

To derive the index, base level scores for each response system were standardized for each condition using the mean and standard deviation appropriate to the response system and condition. These appear in Table 4.4. For each subject, the 12 _z_ scores were then used to form a matrix of ranks, with the largest _z_ score of the three within

---

1. It is possible that the level measure proved most satisfactory for the analysis because of the problems with the response measures noted on p. 61-62. Averaging in the case of bi-directional responses such as RR and HR may have reduced sensitivity of these measures and hence reduced their usefulness for analysis of IRS.

Table 4.4

Means and Standard Deviations Employed in the Skin Conductance, Respiration and Heart Rate Base Level Scores in Relaxation, Tone Series, Count-up, and Mental Arithmetic Conditions

	Skin Conductance		Respiration		Heart Rate	
Condition	Mean	SD	Mean	SD	Mean	SD
Relaxation	8.20	5.77	15.67	3.56	80.74	13.37
Tone Series	9.47	8.46	16.15	3.06	78.12	12.21
Count-up	11.39	8.33	15.77	3.62	79.33	12.87
Mental Arithmetic	12.31	9.21	17.91	3.07	85.04	13.81

a condition receiving a rank of 1 and the smallest a rank of 3. There were thus four sets of ranks for each subject, which were used to compute a Kendall's coefficient of concordance (Siegel, 1956). Where a coefficient equal to or greater than .75, which is significant ($p < .05$) from tables published by Siegel (1956), was obtained, the subject was considered to show individual response specificity. When ranks were totalled across conditions for each system, the system with the lowest total rank score was taken to be the subject's most reactive system. Where total rank scores were equal for one or more systems the subject was classified as a dual responder. Where the coefficient was unity, i.e., the rank order of systems was maintained perfectly from system to system, stereotypy was considered to have been demonstrated. Table 4.5 presents the frequency of specificity and stereotypy observed for each response system.

The final step in deriving the index was to assign a score of 1 to a subject who showed specificity or stereotypy and for whom SC was the most reactive system, and a score of zero in all other instances, and then to weight this in terms of the coefficient of concordance obtained. Thus the maximum possible score was 1, obtained by a subject who showed stereotypy with SC the most reactive system, and the minimum was 0, obtained by all subjects who failed to show specificity or stereotypy with SC as the dominant system. This index was abbreviated as IRSSCB.

Table 4.5

Frequency and Percentage of Subjects Showing Maximum Reactivity, Specificity, and Stereotypy by Response System

Physiological Index	Most Reactive		IRS		Stereotypy	
	f	%	f	%	f	%
SC	49	23.9	33	16.1	13	6.3
HR	75	36.6	56	27.3	21	10.2
RR	58	28.3	47	22.9	20	9.8
Dual Responders	23	11.2	18	8.8	--	----
Total	205	100.0	154	75.1	54	26.3

Self Report Indices

Means, standard deviations, and intercorrelations for scores on the four scales of the EPQ and for scores on the CPI socialization scale are presented in Table 4.6. The data for the EPQ are in close agreement with norms provided by Eysenck and Eysenck (1975) in the test manual. The only substantial discrepancy is the correlation between scores on the lie and neuroticism scales, although the direction of the relationship is consistent with greater defensiveness in the case of the more "neurotic" subjects. Inspection of Table 4.6 indicates that scores on the socialization scale of the CPI correlate with both the neuroticism and psychoticism scores of the EPQ, and in a direction consistent with less socialized subjects having higher scores on neuroticism and psychoticism.

Interrelationship of Individual Difference Measures

Once the various response measures had been analysed separately they were brought together to determine the extent of overlap in the total set of predictors. The intercorrelations are presented in Table 4.7. As both sets of measures of amplitude and level for each condition had shown considerable intercorrelation, aggregate amplitude and level indices were derived by averaging over conditions for each measure. The aggregate amplitude and level indices were abbreviated SCAMP and SCBL respectively. These aggregate indices were used for calculating the intercorrelations shown in Table 4.7. Aggregate indices were also computed for NSR frequency across relaxation and mental arithmetic and for total number of stimulus-evoked responses (TNR) over tone series and count-up.

Discussion

The primary question asked in this first stage of the research concerned the basic dimensions of individual differences in electrodermal responsiveness. A subsidiary question concerned the relationship of these dimensions to the major factors of personality typically assessed by questionnaire. The answer to the first question provided by Podlesny and Raskin (1977) was that three dimensions or aspects of responsiveness can be distinguished: absolute responsiveness of a system, relative responsiveness of the system in comparison to

Table 4.6

Intercorrelations Between Self Report Personality Measures

		E	N	L	P	So[1]
	Mean	13.68	12.40	6.43	3.96	.71
	SD	4.28	4.91	3.69	2.82	.14
E			$-.09^{ns}$	$.02^{ns}$	$-.01^{ns}$	$.07^{ns}$
N				$-.22^{**}$	$.10^{ns}$	$-.34^{***}$
L					$-.19^{**}$	$.15^{*}$
P						$-.39^{***}$
So						

Note: 1. Because a short form of the socialization scale was used with part of the sample, total score was expressed as a ratio of the number of items presented.

*** $p < .001$

** $p < .01$

* $p < .05$

ns not significant

Table 4.7

Means, Standard Deviations and Intercorrelations of Lability, Reactivity, Individual Response Specificity, Personality, and Sex Variables

	SCAMP	SCBL	NSR	TNR	IRSSCB	E	N	L	P	So	Sex
Mean	.94	12.73	8.71	3.67	.17	13.68	12.40	6.43	3.96	.71	.49
SD	2.28	24.79	8.55	2.50	.33	4.28	4.91	3.69	2.82	.14	.50
SCAMP		.97***	.05	.01	.06	-.06	.15*	.09	.17*	-.15*	-.07
SCBL			.06	.00	.14*	-.08	.18*	.00	.20**	-.17*	-.03
NSR				.69***	.32***	-.03	.08	.03	.08	.02	.19**
TNR					.28***	-.06	.00	-.00	.13*	-.01	.25***
IRSSCB						-.18*	.12	.00	.11	-.17*	.19**
E							-.09	.02	-.01	.07	-.07
N								-.22**	.10	-.34***	-.01
L									-.19**	.15*	-.06
P										-.39***	.35***
So											-.26***
Sex											

*** $p < .001$

** $p < .01$

* $p < .05$

others, and amount of change in the system over time. The review of the literature in Chapter 2 provided some support for the distinction between absolute responsiveness and amount of change in the case of electrodermal activity in that a dimension termed lability (amount of change) had been considered by some researchers (e.g., Crider & Lunn, 1971) as orthogonal to absolute responsiveness or reactivity and two factors which could be identifed in this way had emerged in factor analytic research. The literature review also pointed to support for a dimension of relative responsiveness, described as individual response specificity, although the frequency of specificity of electrodermal responsiveness and its relation to the other dimensions were unclear. The results summarized in this chapter confirm the distinction between lability and reactivity but call into question the independence of these dimensions from that of specificity.

Inspection of Tables 4.1, 4.2, and 4.7 indicates the clustering of measures implied by a "two-factor" hypothesis of individual differences in electrodermal responsiveness. Amplitude and level measures form one factor and frequency of nonspecific and evoked responses another. This clustering revealed in Table 4.1 is brought out more sharply in the matrix of factor loadings in Table 4.2, though it is clear from this table that the level and frequency indices conform most clearly to the two factor hypothesis. When aggregate indices were constructed for amplitude, level, and the two types of frequency measure by averaging over the four conditions of the experiment and these aggregate indices intercorrelated (Table 4.7) the conclusion was left in no doubt; aggregate indices for amplitude and level correlate .97, aggregate indices for the frequency measures correlated .69, and the two groups of aggregate indices correlated no more than .06 and in one case zero. Given the sample size on which these correlations are based, the data from this first stage of the research programme would seem to provide strong support for a distinction between electrodermal reactivity and lability.

The terms, reactivity and lability, are of course simply convenient ways of describing the pattern of relationships observed and could be replaced by absolute responsiveness and amount of change, those used by Podlesny and Raskin, without doing violence to the language. This underlines the descriptive nature of the research, in that no greater claim for these factors or dimensions can be made than that they provide convenient summaries of the interrelationship of the variables selected for study. The variables were selected on the basis of previous findings as good examples of what is implied by the terms reactivity and lability, and as a consequence the factors that have emerged can be described in this way. The theoretical meaning of the factors depends on the network of relationships among the measures which define the factors and independent psychological and physiological criteria. For example, whether reactivity reflects individual differences in arousal and lability individual differences in deployment of attention can only be judged by independent studies and cannot be justified on the basis of the correlational and factor analyses reported here.

The conclusions to be drawn with respect to relative responsiveness or specificity suffer much the same limitations as those which apply to conclusions about reactivity and lability. The pattern of results obtained depends on the variables selected for study. Where relative responsiveness is involved the limitation is if anything clearer in that the obvious implication is: responsiveness relative to what? The determination of whether or not a subject is maximally responsive in the electrodermal system depends on which systems are compared. Where not all systems are studied simultaneously, a practical impossibilty if nothing else, it is always possible that a subject described as maximally responsive in the electrodermal system may be shown subsequently to be more responsive in a system which was not monitored. Extending the range of systems monitored reduces but in no way alleviates this problem. The present study, which employed three systems, is vulnerable on this point. Statements about relative responsiveness which follow must therefore be considered to be qualified by the clause: when electrodermal, heart rate, and respiratory responsiveness are studied.

The results for specificity are summarized in Tables 4.5 and 4.7. Inspection of Table 4.5 indicates that the frequency of subjects for whom the electrodermal system is the most reactive (24%) is not high. Although the subjects participating in the present study cannot be considered a representative sample of any specifiable population as systematic sampling techniques were not employed in their recruitment, the size of the sample involved might be considered to give some protection to the inference that maximal responsiveness in the electrodermal system is not a frequent phenomenon. Unfortunately, as the literature review in Chapter 2 indicated, there are few other data available on this point against which the present results might be assessed. If a statistical test is applied in evaluating relative responsiveness, that is some consistency in response patterning is required, then the incidence of relative responsiveness favouring the electrodermal system drops to 16%. Fewer subjects again (6%) showed complete consistency in response patterning in which the electrodermal was the favoured system. These estimates, if substantiated by further systematic sampling, would indicate the size of the problem posed by relative responsiveness in assessment of physiological activity.

For present purposes, however, they point to constraints on the magnitude of the correlations that might be expected between the consistency index and other variables. Relating a dichotomous variable of the sort, responsive/not-responsive, with another involves a restriction on the size of the correlation if the frequencies of cases in the two groups do not equal 50% (Nunnally, 1967). As the frequencies depart from this value, the maximum size of the correlation decreases from its theoretically maximum value of unity. Given that at most 24% of subjects in the present sample comprise the responsive group, a restriction in range in the correlation is to be expected. Partly to offset this problem, an index of relative responsiveness was constructed which took into account the strength of patterning demonstrated in each subject's data (i.e., concordance across conditions). Variance was thereby increased and an index constructed which more faithfully reflected the individual differences of interest.

This index, termed IRSSCB in Table 4.7, correlated significantly with the aggregate indices of frequency of nonspecific and evoked responses and, less strongly, with the aggregate index for the level measures. This finding must be considered to question the distinction between relative responsiveness on the one hand and reactivity and lability on the other. Some overlap between relative responsiveness and certainly lability is indicated in Table 4.7. The overlap is not sufficiently great to warrant interpreting relative responsiveness as simply another expression of lability (or vice versa), but it does indicate that a model of three independent factors to account for individual differences in electrodermal responsiveness is not tenable. Given that the index of relative responsiveness was constructed around individual differences in the level of response, it is interesting to note that the link between this index and the two factors of lability and reactivity is with the former rather than the latter. It is the labile subject who is more likely to be a SC responder. Why this should be so can only be speculated on at this stage, and awaits replication before any attempts are made to interpret it psychologically or physiologically.

Finally, there is the question of the relationship of the dimensions of individual differences in responsiveness to the measures of personality. Measures of the major or "superfactors" (Kline, 1979) of extraversion, neuroticism, and psychoticism were included partly because of the importance attached to these by factor theorists, and partly because previous research on psychophysiological responsiveness had implicated these measures. A measure of socialization was also included because of the findings of Waid et al., with this dimension. The results, however, were largely disappointing.

The measure of extraversion showed no correlation with the measures of electrodermal responsiveness, except for that with relative responsiveness (-.18). That is, more extraverted subjects, those with high scores on the EPQ E scale, were less likely to show maximal responsiveness in SC. The failure to find correlations between extraversion and the lability indices in particular is surprising in view of the conclusions reached by O'Gorman (1977) following his review of the personality correlates of habituation. Although the direction of the correlation between total number of evoked responses, one index of habituation, and extraversion was in the direction expected on the basis of O'Gorman's review, its magnitude was virtually zero. As noted in Chapter 2, however, the evidence on this topic is equivocal.

The measure of neuroticism fared somewhat better in that significant positive correlations, albeit of a low order, with the reactivity indices were observed. The review of the literature by Orlebeke and Feij (1979) on neuroticism and trait anxiety as correlates of amplitude of electrodermal activity pointed to an inconsistent picture. These authors concluded that trait anxiety and amplitude were positively correlated when state anxiety is low but negatively correlated when state anxiety is high. This conclusion was based on a comparison of the pattern of results with psychiatric patients and those with normal controls. The patients were considered to be high in both state and trait anxiety whereas the controls were low in state anxiety but varied in trait anxiety. The present results can be thought of as consistent with the conclusion of Orlebeke and Feij in as much as the subject sample employed would in their terms be low in state anxiety.

The psychoticism dimension, of the three superfactors, showed most correlation with responsiveness. Far less research has been reported to date using this dimension than the other two, and it is therefore somewhat difficult to assess the present findings. James and Barry (1980), however, reported that high P (psychoticism) scorers habituated SC more slowly than low P scorers. Psychoticism, in the present study, was found (Table 4.7) to correlate positively with amplitude, level, and total number of evoked responses, indicating both greater reactivity and lability (though the correlation of number of nonspecific responses was not statistically significant) in the more "psychotic" subject. This result, as far as reactivity is concerned, was not redundant with the findings for neuroticism since the two personality dimensions were, as Eysenck and Eysenck (1975) would maintain, virtually uncorrelated (.10). The implication is that within the normal (i.e., non-clinical) range the more "disturbed" individual, as indicated by high P and N scores, is the more electrodermally reactive.

A similar interpretation can be made with respect to the significant correlations between socialization and the reactivity indices, with the poorly socialized subject being more responsive. Interpretation of this correlation is more difficult, however, as socialization is correlated with both N and P, and the findings with socialization might thus be considered to be redundant with those for the superfactors. It should be noted that contrary to expectation E and socialization were not related.

The other index of individual differences included in Table 4.7 is sex of subject. Males were found to be more labile than females, a result which has some support in the literature (see O'Gorman, 1977), and to be more likely to respond more in the electrodermal system, a result in need of replication. Males were also found to be more "psychotic", consistent with Eysenck and Eysenck's findings with this scale and their neurohumoral interpretation of its biological basis, and less socialized, again a result consistent with previous findings.

In summary, two clear dimensions of individual differences in electrodermal responsiveness emerged from the data gathered in this first stage of the study: lability and reactivity. The factor of relative responsiveness did not appear as separate from these two as anticipated, correlating as it did with the lability indices. These dimensions in turn showed some generally low correlation with the self report tests, the most discernible trend being for the more disturbed subjects to be more reactive and for the less socialized and the male subjects to be more labile.